RANDOM REGULAR DIGRAPHS: SINGULARITY AND DISCREPANCY

NICHOLAS A. COOK

ABSTRACT. We show that the adjacency matrix M of a uniform random d-regular directed graph on n vertices is invertible with high probability, assuming that $d = \lfloor \delta n \rfloor$ for some fixed $\delta \in (0,1)$ and n is large. The proof exploits both local and global symmetries of the distribution of M. As in the analogous work of Komlós for i.i.d. sign matrices, we separately handle the event that M has null vectors with a certain special structure, and employ an anti-concentration estimate for random walks due to Erdős. To overcome difficulties arising from the dependencies among the entries of M we make use of some discrepancy properties for the digraph.

Contents

1. Introduction

For d a positive integer, let $\mathcal{M}_d$ be the set of $n \times n$ matrices with entries taking values in $\{0,1\}$ satisfying the constraint that all row and column sums are equal to d. We let M denote a uniform random element of $\mathcal{M}_d$. One may interpret the elements of $\mathcal{M}_d$ as the adjacency matrices for d-regular digraphs – that is, directed graphs on n vertices where each vertex has exactly d in-neighbors and d out-neighbors (allowing self-loops). One can also identify $\mathcal{M}_d$ with the set of d-regular bipartite graphs with parts of size n in the obvious way. We refer to M as the *random regular digraph* matrix, or "r.r.d. matrix" for short.

We will frequently contrast M with the well-studied *i.i.d. matrices* whose entries are i.i.d. random variables with mean 0 and variance 1. A particular example is the i.i.d. sign

matrix B whose entries are independent ±1-valued Bernoulli random variables. Note that the entries of M are not independent, due to the constraints on row and columns sums.

A basic problem in random matrix theory is to show that the matrix at hand is invertible asymptotically almost surely – that is, that

$$\mathbf{P}(M \text{ is singular}) = o(1)$$

where here and throughout asymptotic notation refers to the large n limit (see Section 1.4 for definitions of asymptotic notation used in this paper). The closely related (and strictly harder) problem of proving that the least singular value $\sigma_n(M)$ is well-separated from zero is a key step in the standard "Hermitization" approach for establishing limit laws for the distribution of eigenvalues of non-Hermitian random matrices – see the survey [9].

Let us begin with a brief review of some known results for i.i.d. matrices. For the invertibility problem for the sign matrix B, in [23] Komlós made use of a bound of Littlewood-Offord type (Erdős' Theorem 2.2 below) to show that

$$\mathbf{P}(B \text{ is singular}) = O(n^{-1/2}).$$

This was subsequently improved to the exponential bound

$$\mathbf{P}(B \text{ is singular}) = O(.999^n), \tag{1.1}$$

by Kahn, Komlós and Szemerédi in [22], where it was also conjectured that the true base of the exponent should be asymptotic to 1/2, that is

$$\mathbf{P}(B \text{ is singular}) = \left(\frac{1}{2} + o(1)\right)^n. \tag{1.2}$$

One verifies that (1.2) is sharp up to the $o(1)$ term by considering the event that the first two columns of B are parallel. The current best bound

$$\mathbf{P}(B \text{ is singular}) \le \left(\frac{1}{\sqrt{2}} + o(1)\right)^n$$

was proved by Bourgain, Vu and Wood in [11], refining the *inverse Littlewood-Offord theory* approach introduced by Tao and Vu in [48].

Many questions in random matrix theory concern the limiting behavior of the singular values

$$\sigma_1(M) \ge \cdots \ge \sigma_n(M) \ge 0$$

and the eigenvalues

$$\lambda_1(M), \dots, \lambda_n(M) \in \mathbf{C}$$

(labeled in some arbitrary fashion). Indeed, the invertibility problem is to bound $\mathbf{P}(\sigma_n(M) = 0)$. The bulk distribution of the singular values of B, or any i.i.d. matrix, is known to be governed for large n by the *quarter-circular law*: letting

$$\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\sqrt{n}}\sigma_i(B)}$$

denote the (rescaled) *empirical singular value distribution* of B, we have that almost surely ν_n converges weakly to the nonrandom measure ν_{QC} supported on $[0, 2]$ with density

$$\frac{1}{\pi}\sqrt{4 - x^2}.$$

See [29] for more details.

Control on the least singular value $\sigma_n(B)$ proved to be a challenging problem. In [38], Rudelson obtained polynomial lower bounds on $\sigma_n(B)$. His estimates were subsequently extended in [41] and [49], using the machinery of inverse Littlewood-Offord theorems developed in [49]. These bounds were an essential ingredient in the proof by Tao and Vu in [51] of the *circular law*, which states that almost surely, the (rescaled) *empirical spectral distribution*

$$\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\frac{1}{\sqrt{n}}\lambda_i(B)}$$

converges weakly to the (nonrandom) uniform measure on the unit disk in $\mathbf{C}$. The circular law was later extended by Wood in [56] to cover sparse i.i.d. Bernoulli matrices, with entries taken to be 0/1-valued $\mathrm{Ber}(p)$ variables, with p allowed to be as small as $n^{\varepsilon-1}$ for any fixed $\varepsilon > 0$. One may interpret such matrices as adjacency matrices of sparse directed Erdős-Rényi graphs.

The above results for the least singular value of B, as well as the quarter circular law and the circular law, are in fact *universal*, in the sense that they hold not only for the i.i.d. sign matrix B, but for any random matrix with independent entries having zero mean and unit variance. See [9] for a survey of works on the circular law for i.i.d. matrices and other matrix models. The assumption of joint independence of the entries has been relaxed in some directions: for instance, in [2] the circular law is established for matrices with log-concave isotropic unconditional laws, and in [3] for matrices with exchangeable entries. While the rows and columns of the r.r.d. matrix M considered in the present work are exchangeable, the individual entries are not.

Outside of i.i.d. matrices, a lot of activity has concentrated on random matrix models with constraints on row and column sums. In [7], Bordenave, Caputo and Chafaï proved the circular law for random Markov matrices, obtained by normalizing the rows of an i.i.d. matrix with continuous entry distributions. More recently, in [8] the same authors have considered Markov generators on sparse random digraphs, proving that the spectrum converges weakly to a certain Gaussian deformation of the circular law. Analogously, the spectrum of Hermitian Markov generators was shown to be governed by the *free convolution* of the Gaussian and semi-circular distributions by Bryc, Dembo and Jiang in [12].

A particularly motivating discrete example is the adjacency matrix of a random undirected regular graph, which is the symmetric analogue of the r.r.d. matrix M. We have the following conjecture of Vu:

Conjecture 1.1. *[55] Fix $d \geq 3$ and let A be the adjacency matrix of a uniform random d-regular (undirected) graph on n vertices (assume nd is even). Then A is invertible asymptotically almost surely, i.e.*

$$\mathbf{P}(A \textit{ is singular}) = o_{n\to\infty}(1).$$

In [33], Nguyen considered a dense, asymmetric version of the above matrix A. Let Q be a random 0/1 matrix with independent rows, where each row is sampled uniformly with the constraint that its entries sum to $n/2$ (assume n is even). It was shown that

$$\mathbf{P}(Q \text{ is singular}) = O_C(n^{-C}).$$

The proof proceeds by what we refer to as the *restriction* strategy, which is a way to borrow results from i.i.d. matrices by viewing the distribution of Q as a (renormalized)

restriction of the product measure on the full space of 0/1 matrices. For instance, suppose we want to control the event that some property P holds for the first row R_1 of Q. Letting $\mathcal{E}$ denote the event that a vector X of n i.i.d. Bernoulli variables sums to $n/2$, we note that

$$\mathbf{P}(\mathcal{E}) = \Omega(n^{-1/2}).$$

Now conditioned on $\mathcal{E}$, X is identically distributed to R_1, so we can deduce the bound

$$\mathbf{P}(P \text{ holds for } R_1) \le \frac{\mathbf{P}(P \text{ holds for } X)}{\mathbf{P}(\mathcal{E})} \ll n^{1/2}\mathbf{P}(P \text{ holds for } X). \tag{1.3}$$

The loss of a factor $n^{1/2}$ is acceptable, reducing the problem to controlling $\mathbf{P}(P \text{ holds for } X)$, which in turn can be handled using the inverse Littlewood-Offord theory developed in [48] and [35] for i.i.d. matrices. Nguyen and Vu subsequently proved the circular law for a more general class of random discrete matrices with constraints on row sums in [36].

The restriction strategy has been applied to other random matrix models with row and column sum constraints. In [54] it was used to deduce that the semicircular law holds on short intervals for the adjacency matrix A of a random regular graphs with degree d growing sufficiently rapidly with n. (For bounded d the empirical spectral distribution is instead governed in the limit by the Kesten-McKay distribution – see [31].) The analogous result for the singular value distribution of rectangular matrices with fixed row and column sums was proved in [53]. See [10], [18] and [16] for similar results obtained using the resolvent method and the fact that sparse regular graphs are locally tree-like.

The strategy has also been applied to matrices with continuous distributions, namely to uniform random doubly stochastic matrices – i.e. matrices sampled uniformly from the normalized volume measure on the Birkhoff polytope. These are in some sense a continuous analogue of the discrete random matrix M considered in this work. In [15], Chatterjee, Diaconis and Sly noted that this measure can be viewed as a restriction of the product exponential measure. This allowed them to deduce results for uniform doubly stochastic matrices that are known for i.i.d. matrices, such as the quarter circular law. Nguyen built on this work in [34] to prove the circular law, again exploiting the connection to i.i.d. exponential matrices.

Let us consider the restriction method for the r.r.d. matrix M. Let B be an $n \times n$ 0/1 matrix with i.i.d. $\mathrm{Ber}(\delta)$ entries for fixed $\delta \in (0,1)$. Let $\mathcal{E}$ denote the event that $B \in \mathcal{M}_d$, with $d = \lfloor \delta n \rfloor$. Conditional of $\mathcal{E}$ we have $B \overset{d}{=} M$. It follows from the work of Canfield and McKay [13] that

$$p_{\delta,n} := \mathbf{P}(\mathcal{E}) \sim C\sqrt{\delta(1-\delta)}n \exp\left(-n \log\left(c\delta(1-\delta)n\right)\right) \tag{1.4}$$

for absolute constants $C, c > 0$. While this is sufficiently large to deduce the quarter-circular law (as in [53]), it is far too small for questions of invertibility.

The reader is encouraged to consult the survey [5] for applications of restriction methods to combinatorial problems outside random matrix theory. In this work we take an alternative approach that works more directly with the distribution of the r.r.d. matrix M. Note that the entries of M do not possess any independence or martingale increment structure; in particular, M does not have the joint independence of rows enjoyed by Q.

As in previous work we seek to access some tools that are available for independent random variables, but we get there using coupling constructions rather than restriction. The general situation is as follows: we are interested in bounding the probability that

some property P holds for a random vector X whose distribution μ may not be a product measure, but which nevertheless possesses many symmetries. We use the symmetries to create a new random vector $\widetilde{X} = \Phi_\omega(X)$, where Φ_ω is a *random* transformation on the range of X such that $\widetilde{X} \overset{d}{=} X$. Now we can work instead with $\widetilde{X}$ and condition on the original variable X:

$$\mathbf{P}(P \text{ holds for } X) = \mathbf{P}(P \text{ holds for } \widetilde{X}) = \mathbf{E}\,\mathbf{P}(P \text{ holds for } \widetilde{X}|X),$$

proceeding to bound $\mathbf{P}(P \text{ holds for } \widetilde{X}|X)$ using only the randomness of Φ_ω. We are in a better position than before if we can somehow define Φ_ω using several independent random variables. This is often possible if the distribution of X has several symmetries of a "local" nature.

Construction of a coupling $(X, \widetilde{X})$ as above is of course a trick that has been exploited by probabilists for some time, most notably in Stein's method for normal approximation [44]. For the present setting of proving invertibility of random matrices, our use of couplings is nicely illustrated by the recent work of Rudelson and Vershynin in [41] on controlling the least singular value of random matrices of the form $D + U$, where D is a deterministic matrix and U is a Haar distributed unitary or orthogonal matrix. For our purposes it will be sufficient to discuss the unitary case. In that work they are able to obtain a bound of the form

$$\mathbf{P}(\sigma_n(D+U) \le t) \le t^c n^C \tag{1.5}$$

for $c, C > 0$ absolute constants, by a method entirely different from the approach that was used for i.i.d. matrices.

We describe the proof of (1.5). Since the rows and columns of U are ℓ_2-normalized we again have the difficulty of dependence among the matrix entries. The strategy is to take advantage of the invariance of the distribution of U under multiplication by other unitary matrices, which can be used to "inject" a lot of independent random variables. Indeed, if L and G are unitary matrices, then $D + U$ is identically distributed to

$$D + LGU.$$

Now we may take L and G to be random unitary matrices independent of each other and of U, with any distribution we like. Upon conditioning on U, we might hope that we can obtain (1.5) using only the randomness of L and G.

The matrix L is taken to be a "local" transformation of the form $I + \varepsilon S$ with S skew-Hermitian; for ε sufficiently small this is close in operator norm to a unitary matrix in a small neighborhood of the identity. S is designed to have several independent normally distributed entries above the diagonal. The "global" transformation G is taken to be a random modulation $\mathrm{diag}(r, 1, \dots, 1)$, where r is uniform on the unit circle in $\mathbf{C}$. It turns out that the randomness injected through S and r is enough to obtain (1.5). As in all works on the least singular value, the proof boils down to the application of a small-ball estimate for an expression involving the matrix entries, which in this case follows quickly from the fact that all of the random variables that we have injected into the problem have bounded density. See [41] for more details.

The control on the least singular value in (1.5) was used in the proof of the Single Ring Theorem for the limiting spectral distribution of certain random matrices with prescribed singular values; see [21, 41]. It was also used in the proof by Basak and Dembo in [6] of the limiting spectral distribution for the sum of a bounded number of independent Haar

unitary or orthogonal matrices. It was conjectured in [9] that the same law should hold for the sum of a bounded number of independent uniform random permutation matrices, which can be viewed as a sparse version of the matrix M considered here.

1.1. **Main result and conjectures.** In this paper we prove that the random regular digraph matrix M is invertible with high probability, with the assumption that d is of linear size in n. Under the graph theoretic interpretation, this assumption means we are considering *dense* regular digraphs.

Theorem 1.2 (Main Theorem)**.** *Let M be drawn uniformly at random from $\mathcal{M}_d$, with $d = \lfloor \delta n \rfloor$ for some $\delta \in (0,1)$. Then*

$$\mathbf{P}(M \textit{ is singular}) \le C_\delta n^{-1/8}$$

for a constant $C_\delta > 0$ depending only on δ.

Remark 1.3. One can easily show that a matrix $M \in \mathcal{M}_d$ is invertible if and only if the "complementary" matrix M' with entries

$$M'(i,j) = 1 - M(i,j)$$

is invertible. Hence, we may and will assume that $\delta \le 1/2$ for the remainder of the paper.

Remark 1.4. A careful examination of the proof reveals that we still have $\mathbf{P}(M \text{ is singular}) = o(1)$ if we allow δ to shrink with n, as long as $\delta \gg n^{-c}$ for a sufficiently small absolute constant $c > 0$. For the sake of exposition we do not carefully track the dependence of implied constants on δ in this paper, and will address sparser matrices in a later work.

We believe that when d is of linear size, the singularity probability is exponentially small, similarly to the bound (1.1) for i.i.d. sign Bernoulli matrices.

Conjecture 1.5. *For fixed $\delta \in (0,1)$, let M be a uniform random element of $\mathcal{M}_{\lfloor \delta n \rfloor}$. Then*

$$\mathbf{P}(M \textit{ is singular}) \le Ce^{-cn}$$

for constants $C, c > 0$ depending only on δ.

We also conjecture that r.r.d. matrices are invertible with high probability for much smaller values of d, paralleling Conjecture 1.1:

Conjecture 1.6. *There are absolute constants $C, c > 0$ such that for any $3 \le d \le n-3$ we have $\mathbf{P}(M \textit{ is singular}) \le Cn^{-c}$.*

When d is bounded, considering the event that two columns of M are parallel shows that we cannot hope for better than a polynomial bound on the singularity probability. M is obviously invertible when $d = 1$ as it is a permutation matrix in this case. As for the case $d = 2$, we observe:

Observation 1.7. *Let M be a uniform random element of $\mathcal{M}_2$. Then M is singular asymptotically almost surely.*

Proof Sketch. One first observes that M is identically distributed to

$$P(I + P_0)$$

where P and P_0 are independent permutation matrices, with P uniform random and P_0 uniform among permutation matrices with 0 diagonal (i.e. P_0 is associated to a uniform random *derangement*). Hence, the probability that M is invertible is equal to the probability that

$$I + P_0$$

is invertible. Now we conjugate by a permutation matrix Q to put P_0 in block diagonal form according to its cycle structure. The resulting block matrix

$$I + Q^{\mathsf{T}} P_0 Q$$

has blocks of the form

$$\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Such a block matrix is invertible if and only if it is of odd dimension. Hence, the probability that M is invertible is equal to the probability that a uniform random derangement decomposes into only odd cycles. The reader may verify that for $\sigma \in \mathrm{Sym}(n)$ a uniform random permutation, we have

$$\mathbf{P}(\sigma \text{ contains only even cycles}) = o(1) \tag{1.6}$$

(for the precise asymptotics of this probability see exercise 5.10 in [43]). The result then follows from the fact that a uniform random permutation is a derangement with probability $\Omega(1)$ – see for instance [58]. □

Remark 1.8. One may similarly show that the sum of two independent and uniformly distributed permutation matrices $P_1 + P_2$ is singular asymptotically almost surely.

1.2. **A word on the general strategy.**

We will make use of several symmetries of the distribution of M, but perhaps the most important is the following: letting

$$\mathbf{I}_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{J}_2 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.7}$$

we can replace a 2×2 minor of M by $\mathbf{I}_2$ if it is $\mathbf{J}_2$ and $\mathbf{J}_2$ if it is $\mathbf{I}_2$ – indeed, note that this preserves the row and column sums. If i_1, i_2 and j_1, j_2 are the row and column indices, respectively, of such a minor, then in the associated digraph G this corresponds to alternating between the following edge configurations at vertices i_1, i_2, j_1, j_2:

where we use solid arrows to depict directed edges, and dashed arrows to indicate places where there is no edge (i.e. "non-edges"). In the random graphs literature this basic strategy is known as the method of "switchings", and has been a successful tool in the study of random regular graphs since its introduction by McKay in [32]; see also section 2.4 of the survey [57].

We can use switchings to "inject" several independent random variables by locating a large number of minors equal to $\mathbf{I}_2$ or $\mathbf{J}_2$ and resampling them independently at random to be equal to either $\mathbf{I}_2$ or $\mathbf{J}_2$. We can encode the outcome of the resampling with i.i.d. Bernoulli variables, which will give us access to an anti-concentration estimate for random walks due to Erdős (Theorem 2.2).

In order for this passage to i.i.d. Bernoulli variables to be useful, much of the proof will be spent ruling out some possible "bad events" stemming from both the dependencies among the entries and the discrete nature of M. It turns out that many of these can be handled by first establishing some *discrepancy properties* for the associated digraph. These ensure that the nonzero entries of M are very uniformly distributed across the matrix, and in particular will guarantee (off a negligibly small event) that there are sufficiently many instances of the minors $\mathbf{I}_2, \mathbf{J}_2$ where we need them. Discrepancy properties for undirected regular graphs similar to those stated in Section 3.1 are already in the literature (see for instance [17, 20] for the sparse case and [26] for the dense case), but we are able to obtain stronger properties for the dense case by a different approach, making use of Chatterjee's method of exchangeable pairs for concentration of measure (see Theorem A.3). The discrepancy properties are proved in the appendix.

To summarize, our approach is to replace M with a coupled matrix $\widetilde{M} = \Phi_\omega(M)$, dividing the randomness into two levels:

(1) the original matrix M, for which our only task is to establish discrepancy properties, and
(2) the independent random variables injected through Φ_ω, which can be used with the discrepancy properties of M to bound the singularity probability via existing anti-concentration estimates.

This general approach may be useful for other discrete random matrices with dependent entries. Following work for invertibility of i.i.d. random matrices, we will divide the set of possible null vectors for M into different classes (this approach goes back to Komlós [23]), and for each class we will use different coupling constructions.

1.3. **Organization of the paper.** The rest of the paper is organized as follows. In Section 2 we describe the ideas of the proof in more detail by comparison to the strategy employed by Komlós for i.i.d. sign matrices. We present there a proposition to rule out the existence of certain structured null vectors for M (Proposition 2.7). Section 3 gathers some important tools that will be used throughout the proof: in Section 3.1 we state the discrepancy properties used in this work, in Section 3.2 we develop the necessary lemmas for injecting independence through switchings, and in Section 3.3 we prove some lemmas for locating minors of the form (1.7) by random sampling. In Section 4 we use these tools and Proposition 2.7 to prove Theorem 1.2. In Section 5 we prove Proposition 2.7. The appendix contains proofs of the discrepancy properties stated in Section 3.1, as well as a deterministic consequence of these properties (Lemma 5.3) that is needed in the proof of Proposition 2.7.

1.4. **Notation.** We consider n as an asymptotic parameter tending to infinity. $f \ll g$, $g \gg f$, $f = O(g)$, and $g = \Omega(f)$ are all synonymous to the statement that $|f| \le Cg$ for all n and for C some absolute constant. $f \asymp g$ means $f \ll g$ and $f \gg g$. $f \ll_p g$, $f = O_p(g)$ etc. mean that $|f| \le C_p g$ for all n, with C_p a constant depending only on the parameter

p. $f = o(g)$ means that $f/g \to 0$ as n tends to infinity. C, c, c', c_1, etc. denote absolute constants whose value may change from line to line. We stress that while many constants are left unspecified, the proof is completely effective.

The factor $\delta(1-\delta)$ will show up repeatedly. We denote it by v_δ ("v" for *variance*), since this is the variance of the entry random variables of M.

Most events will be denoted by the letters $\mathcal{E}, \mathcal{B}$, and $\mathcal{G}$, with various subscripts, where the latter two denote "bad" and "good" events, respectively. Their meaning may vary from proof to proof, but will remain fixed for the duration of each proof. We will also use $\mathcal{L}, \mathcal{R}, \mathcal{S}$ for some additional specific events. $1_\mathcal{E}$ denotes the indicator random variable corresponding to the event $\mathcal{E}$. $\mathbf{E}_X$ and $\mathbf{P}_X$ denote expectation and probability, respectively, conditional on all random variables but X.

We make use of the following terminology for sequences of events.

Definition 1.9. *An event $\mathcal{E}$ depending on n occurs*

- asymptotically almost surely *if* $\mathbf{P}(\mathcal{E}^c) = o(1)$,
- with high probability *if there is a constant $C > 0$ such that* $\mathbf{P}(\mathcal{E}^c) = O(n^{-C})$,
- with overwhelming probability *if* $\mathbf{P}(\mathcal{E}^c) = O_C(n^{-C})$.

An important property of events holding with overwhelming probability is that an intersection of polynomially many of them holds with overwhelming probability.

The variables S and T, with various superscripts and subscripts, will denote subsets of row and column indices, respectively. For $k \le l$ elements of $\mathbf{N}$ we abbreviate $[k] := \{1, \dots, k\}$ and $[k, l] = \{k, \dots, l\}$. $|S|$ denotes the cardinality of a set S.

For a vector $x \in \mathbf{R}^n$, we use notation that views x as a function from $[n]$ to $\mathbf{R}$. In particular, the ith component of x is denoted $x(i)$. We also define the support of x as

$$\mathrm{spt}(x) := \{i : x(i) \neq 0\}.$$

Given a subset $S \subset [n]$, let $(x)_S \in \mathbf{R}^{|S|}$ denote the restriction of x to the components S, retaining the order of the components. We let $\mathbf{1}$ denote the all-ones vector $(1, \dots, 1) \in \mathbf{R}^n$.

We will frequently need to refer to a specific minor of M. Given ordered tuples of row and column indices $(i_1, \dots, i_s)$ and $(j_1, \dots, j_t)$, we denote by $M_{(i_1,\dots,i_s)\times(j_1,\dots,j_t)}$ the $s \times t$ matrix with (k, l) entry equal to the (i_k, j_l) entry of M. (Note for instance that the sequence $(i_1, \dots, i_s)$ need not be increasing.) When we write $M_{S\times T}$ with S and T unordered subsets of $[n]$, the natural ordering of $\mathbf{N}$ is implied. For example, we denote the top-left and bottom-right $k \times k$ minors of M by $M_{[k]\times[k]}$ and $M_{[n-k+1,n]\times[n-k+1,n]}$, respectively.

For the special case of 2×2 minors, we use $\mathbf{p}, \mathbf{q}$ to label ordered pairs of row and column indices, respectively. The rows and columns of M will simply be denoted by R_i, X_j, respectively.

"Null vector" will mean "right null vector" unless otherwise stated.

1.5. **Acknowledgements.** The author thanks Terence Tao for invaluable discussions on this topic and random matrix theory in general, and for helpful feedback on the manuscript.

2. Ideas of the proof

The approach to Theorem 1.2 is inspired by Komlós' argument for an analogous theorem about i.i.d. sign matrices. We review Komlós' proof below and discuss the new ideas that are necessary to deal with the r.r.d. matrix M.

Theorem 2.1 (Komlós [23])**.** *Let B be an $n \times n$ i.i.d. sign matrix. Then*

$$\mathbf{P}(B \textit{ is singular}) = O(n^{-1/2}).$$

The event $\mathcal{S}$ that B is singular can be written

$$\mathcal{S} = \{\exists \text{ nontrivial } x \in \mathbf{R}^n : Bx = 0\}.$$

Komlós introduces a general strategy to control the singularity event $\mathcal{S}$ by separately ruling out the existence of structured and non-structured null vectors x. For the above theorem the right notion of structure is *sparsity*; indeed, the key insight of Komlós is to note that the existence of *non-sparse* null vectors can be handled using the following anti-concentration estimate for random walks due to Erdős.

Theorem 2.2 (Erdős [19])**.** *Let $x \in \mathbf{R}^n$ have at least k nonzero components, and let $\xi = (\xi_1, \dots, \xi_n)$ be a vector of i.i.d.* $\mathrm{Ber}_{\pm}(1/2)$ *random variables. Then*

$$\sup_{a \in \mathbf{R}} \mathbf{P}(x \cdot \xi = a) = O(k^{-1/2}). \tag{2.1}$$

Proof of Theorem 2.1. We will roughly follow the exposition in [45]. Say $x \in \mathbf{R}^n$ is k-sparse if $|\operatorname{spt}(x)| \le k$. It turns out that the event that B has a sparse null vector is very small:

Proposition 2.3 (No structured null vectors for B)**.** *For any fixed $\varepsilon \in (0,1)$, with overwhelming probability B has no nontrivial $(1-\varepsilon)n$-sparse null vectors.*

Remark 2.4. We will observe in the proof that the conclusion still holds for ε as small as $Kn^{-1/4}$ for some K sufficiently large. This is not needed to prove Theorem 2.1, but a similar observation for the case of r.r.d. matrices will be useful – see Proposition 2.6.

We save the proof of this proposition for the end. Note that since B is identically distributed to its transpose, we have the same bound for the case of sparse *left* null vectors. Fix $\varepsilon \in (0,1)$. It remains to control the event $\mathcal{S}'$ that B has a null vector but has no $(1-\varepsilon)n$-sparse left or right null vectors.

For each $i \in [n]$, let X_i denote the ith column of B and let

$$V_i = \operatorname{span}(X_j : j \neq i).$$

Denote the events

$$\mathcal{S}_i := \mathcal{S}' \wedge \{X_i \in V_i\}.$$

Now on $\mathcal{S}'$, the existence of a non-εn-sparse (right) null vector implies that $\mathcal{S}_i$ holds for at least $(1-\varepsilon)n$ values of i. By double counting we then have that

$$\sum_{i=1}^n \mathbf{P}(\mathcal{S}_i) \ge (1-\varepsilon)n\, \mathbf{P}(\mathcal{S}').$$

By column exchangeability, all of the summands on the left hand side are equal to $\mathbf{P}(\mathcal{S}_1)$, so we conclude

$$\mathbf{P}(\mathcal{S}') \leq \frac{1}{1-\varepsilon}\,\mathbf{P}(\mathcal{S}_1). \tag{2.2}$$

It remains to control the event $\mathcal{S}_1$.

We condition on the columns $X_2, \ldots, X_n$, which fixes their span V_1. We now select a unit normal u of V_1 arbitrarily (but independently of X_1), say uniformly at random. Now the event that $X_1 \in V_1$ is contained in the event that X_1 is perpendicular to u. To summarize:

$$\begin{aligned}\mathbf{P}(\mathcal{S}_1) &= \mathbf{E}\,\mathbf{P}(\mathcal{S}_1 | X_2, \ldots, X_n) \\ &\leq \mathbf{E}\,\mathbf{P}\big(\mathcal{S}' \wedge \{X_1 \cdot u = 0\} | X_2, \ldots, X_n\big).\end{aligned}$$

Now we note that on the event $\mathcal{S}' \wedge \{X_1 \cdot u = 0\}$, u is not $(1-\varepsilon)n$-sparse. Indeed, if it were, then u would be perpendicular to all of the columns of B, and hence a left null vector, putting us in the complement of $\mathcal{S}'$. By independence, the conditioning on the columns $X_2, \ldots, X_n$ has not affected the distribution of X_1, so we can apply Theorem 2.2 to the random walk $X_1 \cdot u$ to conclude

$$\mathbf{P}(\mathcal{S}_1) \ll_\varepsilon n^{-1/2}.$$

Combining this with (2.2) completes the proof, on Proposition 2.3.

We turn to Proposition 2.3. Define the events

$$\mathcal{E}_k = \{\exists \text{ nontrivial } x \in \mathbf{R}^n : x \text{ is } k\text{-sparse, } Bx = 0\}.$$

Since

$$\mathbf{P}(\mathcal{E}_{(1-\varepsilon)n}) = \sum_{k=2}^{\lfloor (1-\varepsilon)n \rfloor} \mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \tag{2.3}$$

we will bound $\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ for each $k \in [2, (1-\varepsilon)n]$.

Fix a k in this range. On $\mathcal{E}_k \setminus \mathcal{E}_{k-1}$ there is a right null vector x with exactly k nonzero components. We may spend a factor of $\binom{n}{k}$ to assume that x is supported on $[k]$. Now since we are on the complement of $\mathcal{E}_{k-1}$, we know that the first k columns of B span a space of dimension $k-1$, from which it follows that there are $k-1$ linearly independent rows of the minor $M_{[n]\times[k]}$. Spending another factor $\binom{n}{k-1}$ to assume that the first $k-1$ rows are linearly independent, we have

$$\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq \binom{n}{k}\binom{n}{k-1}\mathbf{P}(\mathcal{E}'_k) \tag{2.4}$$

where

$$\mathcal{E}'_k = \{\exists x \in \mathbf{R}^n : Bx = 0,\ \ \mathrm{spt}(x) = [k],\ R_1, \ldots, R_{k-1} \text{ linearly independent}\}.$$

Note that by linear independence, $R_1, \ldots, R_{k-1}$ determine x. Conditioning on these rows fixes x, and we are left with

$$\mathbf{P}(R_i \cdot x = 0\ \forall k \leq i \leq n) = \mathbf{P}(R_n \cdot x = 0)^{n-k+1} \tag{2.5}$$

where we have used the independence of the rows of B. By (2.1), this is bounded by $O(k^{-1/2})^{n-k+1}$ (for small values of k we may use the crude bound $\Pr(R_n \cdot x = 0) \leq 1/2$ in

place of Erdős' bound). Combined with (2.4) and (2.3) we conclude using the inequality $\binom{n}{k} \le (en/k)^k$ that

$$\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \ll \exp\left((n-k)\left(C + 2\log\frac{n}{n-k} - \log\sqrt{k}\right)\right) \tag{2.6}$$

Summing these over k we find

$$\mathbf{P}(\mathcal{E}_{(1-\varepsilon)n}) = O_\varepsilon(\exp(-c_\varepsilon n)) \tag{2.7}$$

as desired. The reader may verify from (2.6) that we can actually take $\varepsilon = o(1)$ as long as $\varepsilon \ge Kn^{-1/4}$ for K a sufficiently large constant. □

We now discuss the difficulties one encounters in trying to apply the above argument to the r.r.d. matrix M. A key element of the proof was the reduction to the event that a single column X_1 was in the span V_1 of the remaining columns. We were then able to control this event by conditioning on the columns $X_2, \dots, X_n$ and selecting a unit normal vector of V_1. However, for the r.r.d. matrix M, conditioning on $n-1$ columns also fixes the remaining column due to the row-sums constraint, so we will need to leave at least two columns free.

As a warmup let us try to control the event $\mathcal{S}_{1,2}$ that the first two columns X_1, X_2 of M lie in the span $V_{1,2}$ of the remaining columns. We will see later (Lemma 4.1) that by a double counting argument similar to the one in the proof of Theorem 2.1 above, control on the event $\mathcal{S}_{1,2}$ gives control on the event that M is of co-rank at least 2.

Following Komlós' argument, we condition on the columns $X_3, \dots, X_n$ and seek control on the event that both X_1 and X_2 are orthogonal to a vector y in $V_{1,2}^\perp$:

$$\left\{ \begin{pmatrix} X_1^\mathsf{T} \\ X_2^\mathsf{T} \end{pmatrix} y = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}. \tag{2.8}$$

In the proof of Theorem 2.1 this was accomplished by interpreting the dot product $X_1^\mathsf{T} y$ as a random walk and appealing to Erdős' estimate from Theorem 2.2, after arguing that y must be non-sparse. However, at first glance it seems that Erdős' estimate is of no use here, as the entries of X_1, X_2 are not i.i.d. Bernoulli variables.

Nevertheless, we can "inject" Bernoulli variables into the joint distribution of X_1, X_2 in the following way. We randomly sample ordered pairs of row indices $\{(I_l^+, I_l^-)\}_{l=1}^m$ from $[n]$; we do not discuss here exactly how these indices are sampled, but we assume that all $2m$ of them are distinct. For each $l \in [m]$, we look at the randomly sampled 2×2 minor

$$M_{(I_l^+, I_l^-)\times(1,2)}. \tag{2.9}$$

We say that this minor is "switchable" if it is equal to either

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

If the minor is switchable, we randomly resample it to be equal to either $\mathbf{I}_2$ or $\mathbf{J}_2$, uniformly and independently of all other switchings. We encode the random resampling with a Bernoulli variable ξ_l, equal to $+1$ it the minor is resampled as $\mathbf{I}_2$ and -1 if it is resampled as $\mathbf{J}_2$. If the minor is not switchable, we leave it unchanged. We let $\widetilde{M}$ denote the matrix obtained after performing the random switchings. Note that $\widetilde{M}$ and M differ only on a random subset of the entries of the first two columns. See Figure 1.

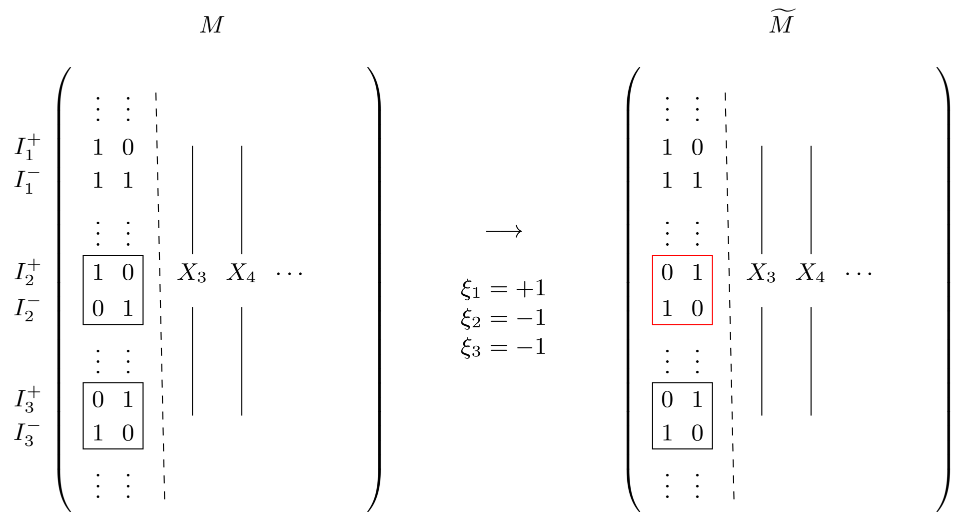

FIGURE 1. Construction of a coupled pair of r.r.d. matrices $(M, \widetilde{M})$. 2×2 minors of the first two columns of M are sampled by selecting pairs of row indices $\{(I_l^+, I_l^-)\}_{l=1}^m$ at random. For clarity we have depicted I_l^+, I_l^- next to each other for $l = 1, 2, 3$, though this need not be the case. The sampled minors $A_l = M_{(I_l^+, I_l^-) \times (1,2)}$ are

$$A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The latter two are switchable and are resampled according to ξ_2, ξ_3. Both of these are -1, which means to replace A_2, A_3 with $\mathbf{J}_2$. ξ_1 is ignored as A_1 is not switchable. Hence, out of the three sampled minors the only alteration was on A_2.

In Section 3.2 we will show that the random sampling of 2×2 minors can be done in a way such that the resulting matrix $\widetilde{M}$ is also uniformly distributed in $\mathcal{M}_d$. We are hence free to replace M with $\widetilde{M}$ in (2.8). Conditioning on M and the sampled pairs $\{(I_l^+, I_l^-)\}_{l=1}^m$, the dot product $X_1^\mathsf{T} y$ is now a random walk:

$$X_1^\mathsf{T} y = \hat{X}_1^\mathsf{T} y + \sum_{l \in L} X_1(I_l^+) y(I_l^+) + X_1(I_l^-) y(I_l^-) \tag{2.10}$$

$$= \hat{X}_1^\mathsf{T} y + \sum_{l \in L} a_l(y) + \xi_l \partial_l(y) \tag{2.11}$$

where L is the set of $l \in [m]$ for which the minor $M_{(I_1^+, I_1^-)}$ is switchable, $\hat{X}_1$ is the restriction of X_1 to the components that are unaltered by the switchings, and we have set

$$a_l(y) := \frac{y(I_l^+) + y(I_l^-)}{2}$$

and

$$\partial_l(y) := \frac{y(I_l^+) - y(I_l^-)}{2}.$$

With the conditioning on M and $\{(I_l^+, I_l^-)\}_{l=1}^m$ the only randomness is in the $\mathrm{Ber}_{\pm}(1/2)$ variables $\{\xi_l\}_{l\in L}$. Looking at the coefficient $\partial_l(y)$ of ξ_l in (2.11), we see that if $y(I_l^+) \neq y(I_l^-)$ for a large number of $l \in [m]$, then Erdős' estimate (2.1) can be applied to get a good bound on the probability that $X_1^{\mathsf{T}} y = 0$. We hence see that while Komlós had to address the case that the span of $n-1$ columns had a sparse normal vector, we will have to deal with the possibility that the span of $n-2$ columns has a normal vector with many pairs of identical components.

Furthermore, some problems may arise from the dependencies due to the row sum condition. By conditioning on the columns $X_3, \ldots, X_n$, we have fixed the rows R_i whose first two entries are $(1, 0)$ or $(0, 1)$, since these are the rows for which

$$\sum_{j=3}^{n} R_i(j) = d - 1.$$

The columns $X_3, \ldots, X_n$ hence influence the location of switchable 2×2 minors. Let y denote a vector normal to the span of $X_3, \ldots, X_n$, and let S be the set of components i where $M_{i\times(1,2)}$ is $(1 \quad 0)$ or $(0 \quad 1)$. It is possible that $X_3, \ldots, X_n$ will determine y and S in such a manner that y is mostly constant on S. In this case the randomness of the ξ_l alone will not be enough to control the event that $X_1^{\mathsf{T}} y = 0$, since most of the steps $\partial_l(y)$ of the random walk (2.11) will be zero.

It will hence be necessary to use the randomness of M and the sampling $\{(I_l^+, I_l^-)\}_{l=1}^m$ to argue that with high probability, for any $y \in V_{1,2}^{\perp}$ we will have $y(I_l^+) \neq y(I_l^-)$ for many values of l. We do this in two steps.

First, we show that in the randomness of M, two properties hold with overwhelming probability:

(1) *Discrepancy property*: the number of indices $i \in [n]$ such that

$$M_{i\times(1,2)} = (1 \quad 0)$$

is close to its expected value, which is roughly $\delta(1-\delta)n$, and similarly for $(0 \quad 1)$. (See Proposition 3.1.)

(2) *No structured null vectors*: any vector $y \in V_{1,2}^{\perp}$ has *small level sets*, which is to say that the proportion of components of y taking any fixed value is $o(1)$ (see Definition 2.5 below).

Second, conditioning on M having the above two properties, we can show that in the randomness of the sampling $\{(I_l^+, I_l^-)\}_{l=1}^m$, with overwhelming probability the number of $l \in [m]$ such that

$$M_{(I_l^+, I_l^-)\times(1,2)} \text{ is switchable} \quad \textit{and} \quad \partial_l(y) \neq 0$$

is $\Omega_\delta(m)$. Finally, conditioning also on this good sampling, we can use the randomness of the ξ_l variables with Theorem 2.2 to get a good bound on $\mathbf{P}(\mathcal{S}_{1,2})$.

The above was a sketch of our approach to showing that M has co-rank at most 1 with high probability. To control the event that M has co-rank 1, we show below that it suffices to control the event that

$$\begin{vmatrix} X_1 \cdot u_1 & X_1 \cdot u_2 \\ X_J \cdot u_1 & X_J \cdot u_2 \end{vmatrix} = 0 \tag{2.12}$$

where J is a uniform random index in $[2,n]$ and $u_1 \perp u_2$ are randomly chosen from $V_{1,J}^{\perp}$. Conditional on J, we construct a coupling $(M,\widetilde{M})$ as above and express this determinant as a random walk. It turns out that the randomness of M, the sampled pairs $\{(I_l^+,I_l^-)\}_{l=1}^m$, and the index J are enough to ensure that with high probability there are enough nonzero steps in the random walk to effectively apply the estimate (2.1). See Section 4 for details.

Section 3.1 states the discrepancy properties that hold for M with overwhelming probability; the proofs are deferred to the appendix. Apart from the discrepancy property for a pair of columns discussed above, we will need an "edge discrepancy" property: any sufficiently large minor of M will have roughly a proportion δ of its entries equal to 1 (see Theorem 3.5). Under the digraph interpretation, this says that for sufficiently large sets S,T of vertices, the number of edges passing from S to T is close to $\delta|S||T|$. Discrepancy properties such as these are easily deduced for Erdős-Rényi graphs (or digraphs) with Chernoff-type bounds due to the independence of the edges. Because the entries of M are dependent we will have to take a longer route, making use of Chatterjee's method of exchangeable pairs for concentration of measure.

Now we discuss the "no structured null vectors" property more precisely. Viewing a vector $x\in\mathbf{R}^n$ as a function $x:[n]\to\mathbf{R}$, we may define the *level sets* of x

$$x^{-1}(\lambda):=\{i\in[n]:x(i)=\lambda\}.$$

Definition 2.5. *We say that a vector $x\in\mathbf{R}^n$ has the* small level sets property with parameter ε, *abbreviated* SLS(ε), *if for any $\lambda\in\mathbf{R}$ we have that $|x^{-1}(\lambda)|\le\varepsilon n$.*

The following proposition is the r.r.d. analogue of Proposition 2.3.

Proposition 2.6 (No structured null vectors for M)**.** *With overwhelming probability, any nontrivial null vector of M has* $\mathrm{SLS}(K_\delta n^{-1/8})$ *for some K_δ sufficiently large depending only on δ.*

For technical reasons, at one point we will need the stronger statement that this holds for (right) null vectors of a matrix obtained by removing a small number of rows from M.

Proposition 2.7 (No structured null vectors for M_S)**.** *For $S\subset[n]$, let M_S denote an $(n-|S|)\times n$ matrix obtained from M by deleting the rows indexed by S. For any $r\ge 0$ there is a constant $K_{\delta,r}>0$ depending only on δ,r such that for any fixed $S\subset[n]$ with $|S|=r$, with overwhelming probability any nontrivial right null vector of M_S has* $\mathrm{SLS}(K_{\delta,r}n^{-1/8})$.

We prove these propositions in Section 5 as consequences of a stronger statement (Proposition 5.1) allowing for the removal of up to $\Omega(n)$ rows, but there is a trade-off between the number of rows removed and the small level sets parameter ε. Although we will not need Proposition 5.1 in the present work, we include it because the proof requires essentially no additional effort than is required to prove Proposition 2.7, and because it may be useful in future work.

An immediate consequence of Proposition 2.6 is that M is of co-rank $o(n)$ with overwhelming probability.

Corollary 2.8. *With overwhelming probability we have* $\mathrm{rank}(M)\ge n-K_\delta n^{-1/8}$, *with K_δ as in Proposition 2.6.*

Proof. Suppose $\mathrm{rank}(M) < n - K_\delta n^{-1/8}$. Then for some $j \in [n - K_\delta n^{-1/8}]$ we have that the jth column of M is in the span of the first $j-1$ columns. This defines an $n - K_\delta n^{-1/8}$-sparse linear dependency among the columns, contradicting Proposition 2.6. ☐

3. Preliminary tools

3.1. **Discrepancy properties.** In this section we state certain discrepancy properties that hold with overwhelming probability for the r.r.d. matrix M, through its association to a uniform random d-regular digraph G. We defer the proofs to the appendix. Roughly speaking, G has a discrepancy property if its edges are uniformly spread in some sense.

While analogous results for Erdős-Rényi graphs are easily obtained from Chernoff-type bounds, the present setting of random regular graphs is more difficult due to the lack of independence of the edges. Some discrepancy properties along the lines of Proposition 3.1 and Theorem 3.5 are already in the literature. For instance, a discrepancy property for sparse random regular graphs was a crucial ingredient in the proof by Kahn and Szemerédi in [20] of a bound for the second eigenvalue of the adjacency matrix (a more detailed exposition of this argument is available in [17]). Also, [26] contains weaker versions of Proposition 3.1 and Theorem 3.5 for dense undirected regular graphs (though they are valid for sparser matrices than we consider here). We will comment further on these related results below.

In this paper we use discrepancy properties for codegrees and for edge counts. To define these notions, for now we let $G = (V, E)$ denote an arbitrary directed graph with vertex set $V = [n]$ and edge set $E \subset [n] \times [n]$, and we denote by M the associated (non-random) $n \times n$ adjacency matrix, i.e.

$$M(i,j) = 1_{(i,j)\in E}. \tag{3.1}$$

For a vertex $i \in [n]$, let

$$\mathcal{N}^{out}(i) = \{j \in [n] : (i,j) \in E\}$$

be the set of outgoing neighbors of i, and similarly define

$$\mathcal{N}^{in}(i) = \{j \in [n] : (j,i) \in E\}$$

to be the set of incoming neighbors i. For a pair of vertices $i_1, i_2 \in V$, we define their out-codegree to be

$$\begin{aligned}\mathrm{codeg}_{(i_1,i_2)}(M) &= |\mathcal{N}^{out}(i_1) \cap \mathcal{N}^{out}(i_2)| \\ &= \left|\left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix}1\\1\end{pmatrix}\right\}\right|\end{aligned}$$

and their in-codegree

$$\begin{aligned}\mathrm{codeg}_{(i_1,i_2)}(M^{\mathsf{T}}) &= |\mathcal{N}^{in}(i_1) \cap \mathcal{N}^{in}(i_2)| \\ &= \left|\left\{ j \in [n] : M_{j\times(i_1,i_2)} = (1 \quad 1)\right\}\right|.\end{aligned}$$

Let

$$\overline{\mathrm{codeg}}(M) = \frac{1}{\binom{n}{2}} \sum_{1\le i_1<i_2\le n} \mathrm{codeg}_{(i_1,i_2)}(M).$$

We informally say that M has a *codegree discrepancy property* if the in- and out- codegrees of (i_1, i_2) are close to $\overline{\mathrm{codeg}}(M)$ and $\overline{\mathrm{codeg}}(M^T)$, respectively, uniformly in (i_1, i_2).

It will actually be more convenient for us to phrase codegree discrepancy in terms of the statistics

$$h_{(i_1,i_2)}(M) := \left|\left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}\right|. \tag{3.2}$$

For d-regular digraphs, these statistics have the same information as the codegrees, since the row sum constraint for row i_1 implies

$$\mathrm{codeg}_{(i_1,i_2)}(M) = d - h_{(i_1,i_2)}(M)$$

for any ordered pair of distinct indices (i_1, i_2). One can easily verify that

$$\begin{aligned} \hbar(M) &:= \frac{1}{n(n-1)} \sum_{i_1 \neq i_2 \in [n]} h_{(i_1,i_2)}(M) \\ &= \frac{d(n-d)n}{n(n-1)} \\ &= \delta(1-\delta)n + O(1) \\ &= v_\delta n + O(1). \end{aligned}$$

Proposition 3.1 below states that for the r.r.d. matrix M, $h_{(i_1,i_2)}(M)$ and $h_{(i_1,i_2)}(M^\mathsf{T})$ are uniformly close to $v_\delta n$ with overwhelming probability.

Proposition 3.1 (Codegree discrepancy)**.** *For $\lambda \in (0,1)$ let*

$$\mathcal{B}_\lambda = \bigvee_{\substack{(i_1,i_2) \in [n]\times[n], \\ i_1 \neq i_2}} \left\{ \left| \frac{h_{(i_1,i_2)}(M)}{v_\delta n} - 1 \right| \geq \lambda \right\} \vee \left\{ \left| \frac{h_{(i_1,i_2)}(M^\mathsf{T})}{v_\delta n} - 1 \right| \geq \lambda \right\}$$

where $v_\delta = \delta(1-\delta)$. Then

$$\mathbf{P}(\mathcal{B}_\lambda) \ll n^2 \exp\left(-c v_\delta^2 \min(\lambda, v_\delta) n\right). \tag{3.3}$$

Remark 3.2. An analogue of the above proposition for sufficiently dense *undirected* regular graphs was proved in [26], though they only needed a polynomially small bound on $\mathbf{P}(\mathcal{B}_\lambda)$.

As discussed in Section 2, the above proposition is a fundamental ingredient of our approach to the r.r.d. matrix M, as it allows us to locate several switchable 2×2 minors (in the sense of (1.7)) in a pair of rows or columns. It also serves as a stepping stone to proving edge discrepancy properties, which we now discuss.

Again consider an arbitrary deterministic digraph $G = ([n], E)$ and its adjacency matrix M. For a pair of sets $S, T \subset [n]$, we let

$$e_{S,T}(M) = |\{(i,j) \in E \cap S \times T\}| = \sum_{(i,j) \in S \times T} M(i,j)$$

denote the number of edges passing from S to T. Let

$$\overline{e} = \frac{1}{|V|^2} e_{V,V}(M) = \frac{1}{n^2} \sum_{i,j=1}^{n} M(i,j)$$

be the average edge density of G. If $\mathcal{F} \subset \{(S,T) : S, T \subset [n]\}$ is a family containing pairs of all "sufficiently large" sets of vertices, we may define the *edge discrepancy of M*

at scales $\mathcal{F}$ to be

$$\mathrm{discrep}_{\mathcal{F}}(M) := \sup_{(S,T)\in\mathcal{F}} \left| \frac{e_{S,T}(M)}{|S||T|} - \overline{e} \right|. \tag{3.4}$$

We (somewhat imprecisely) say that G (and hence M) has an *edge discrepancy property* if it has low edge discrepancy for some family $\mathcal{F}$ including all pairs (S,T) above some reasonably fine scale.

For a d-regular digraph G on vertex set $V = [n]$ we have

$$\overline{e} = d/n = \delta + O(1/n).$$

In terms of the associated adjacency matrix M, if G has low edge discrepancy, then all sufficiently large minors $M_{S\times T}$ have close to $\delta|S||T|$ entries equal to 1.

Before stating our result on edge discrepancy for M, we first note that a crude version already follows deterministically from the codegree discrepancy property by an argument of Alon, Krivelevich and Sudakov from [4]. We have the following deterministic lemma:

Lemma 3.3. *For $M \in \mathcal{M}_{\lfloor \delta n \rfloor}$ deterministic, suppose we have*

$$\mathrm{codeg}_{(i_1,i_2)}(M) - \delta^2 n \le \varepsilon n \tag{3.5}$$

for all $i_1, i_2 \in [n]$, $i_1 \neq i_2$. Then for any $\eta \in (0,\delta)$ and any pair of sets $S, T \subset [n]$ with

$$|S||T| \ge \frac{v_\delta + \varepsilon|S|}{\eta^2} n \tag{3.6}$$

we have

$$\big| e_{S,T}(M) - \delta|S||T| \big| \le \eta |S||T|. \tag{3.7}$$

Proof. We follow the lines of [4] (they argued for undirected graphs, but the proof applies just as well to the directed case). Fix $S, T \subset [n]$ as above. From the Cauchy-Schwarz inequality we have

$$\begin{aligned}
(e_{S,T}(M) - \delta|S||T|)^2 &= \Big(\sum_{i\in S}\sum_{j\in T} (M(i,j) - \delta) \Big)^2 \\
&\le |T| \sum_{j\in[n]} \Big(\sum_{i\in S} (M(i,j) - \delta) \Big)^2 \\
&= |T| \sum_{j\in[n]} \Big(\sum_{i\in S} (M(i,j)-\delta)^2 + \sum_{i_1,i_2\in S, i_1\neq i_2} (M(i_1,j)-\delta)(M(i_2,j)-\delta) \Big).
\end{aligned}$$

Switching the order of summation gives

$$\begin{aligned}
(e_{S,T}(M) - \delta|S||T|)^2 &\le |T| \sum_{i\in S}\sum_{j\in[n]} M(i,j)(1-2\delta) + \delta^2 \\
&\quad + |T| \sum_{i_1,i_2\in S, i_1\neq i_2} \sum_{j\in[n]} M(i_1,j)M(i_2,j) - \delta(M(i_1,j) + M(i_2,j)) + \delta^2 \\
&= \delta(1-\delta)|S||T|n + |T| \sum_{i_1,i_2\in S, i_1\neq i_2} \mathrm{codeg}_{(i_1,i_2)}(M) - \delta^2 n \\
&\le \delta(1-\delta)|S||T|n + |S|(|S|-1)|T|\varepsilon n.
\end{aligned}$$

where in the last line we have used (3.5). It follows that

$$\big| e_{S,T}(M) - \delta|S||T| \big| \le \big[(v_\delta + \varepsilon|S|)|S||T|n \big]^{1/2}. \tag{3.8}$$

Now if S, T satisfy (3.6) we conclude

$$\big|e_{S,T}(M) - \delta|S||T|\big| \le \eta|S||T| \tag{3.9}$$

as desired. $\square$

Corollary 3.4 (Edge discrepancy for large minors)**.** *For any $\lambda, \lambda_1 \in (0,1)$, with probability $1 - O\big(n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta)n\right)\big)$, we have*

$$\left|\frac{e_{S,T}(M)}{\delta|S||T|} - 1\right| \le \lambda_1 \tag{3.10}$$

for any pair of sets $S, T \subset [n]$ such that

$$|S||T| \ge \frac{1 + \lambda \min(|S|, |T|)}{\lambda_1^2 \delta} n. \tag{3.11}$$

Proof. Letting $\mathcal{B}_\lambda$ be as in Proposition 3.1, we have that on the complement of $\mathcal{B}_\lambda$,

$$\max\left(\operatorname{codeg}_{(i_1,i_2)}(M), \operatorname{codeg}_{(i_1,i_2)}(M^{\mathsf{T}})\right) - \delta^2 n \le \lambda v_\delta n.$$

Hence, both M and M^{T} satisfy (3.5) with $\varepsilon = \lambda v_\delta$. Let $S, T \subset [n]$ satisfy (3.11), take $\eta = \lambda_1 \delta$, and apply Lemma 3.3 for M if $|S| \le |T|$ and for M^{T} if $|S| \ge |T|$. $\square$

In the present work we will only make use of edge discrepancy with $\lambda_1 \gg_\delta 1$ and for pairs of sets S, T such that

$$\min(|S|, |T|) \gg_\delta n.$$

Hence, Corollary 3.4 suffices for our purposes if we take λ sufficiently small depending on δ. However, with Proposition 3.1 in place it does not take much additional work to establish the following, which gives edge discrepancy for much smaller minors of M.

Theorem 3.5 (Edge discrepancy)**.** *For $K > 0$, define the "coarse-scale" family of pairs of subsets of $[n]$*

$$\mathcal{F}_{\mathfrak{c}}(K) = \{(S,T) : |S||T| \ge Kn\} \tag{3.12}$$

and the larger "fine-scale" family

$$\mathcal{F}_{\mathfrak{f}}(K) = \left\{(S,T) : |S||T| \ge Kn\, G\left(\frac{\max(S,T)}{n}\right)\right\} \tag{3.13}$$

where $G(x) := -x \log x$. For $\lambda \in (0,1)$ and $\mathcal{F}$ a family of pairs of subsets of $[n]$, define the bad event

$$\mathcal{B}_{\lambda,\mathcal{F}} := \left\{\exists\, (S,T) \in \mathcal{F} : \ \left|\frac{e_{S,T}(M)}{\delta|S||T|} - 1\right| \ge \lambda\right\}. \tag{3.14}$$

For any $\lambda \in (0,1)$ there exists $K_{\lambda,\delta} > 0$ depending only on λ, δ such that

$$\mathbf{P}\Big(\mathcal{B}_{\lambda, \mathcal{F}_{\mathfrak{c}}(K_{\lambda,\delta})}\Big) \ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta)n\right) \tag{3.15}$$

and

$$\mathbf{P}\Big(\mathcal{B}_{\lambda, \mathcal{F}_{\mathfrak{f}}(K_{\lambda,\delta})}\Big) \ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta)n\right) + \exp\left(-c\log^2 n\right). \tag{3.16}$$

In fact we can take $K_{\lambda,\delta} = \frac{C}{v_\delta^3 \lambda^2 (1-\lambda)^2}$ for a sufficiently large constant C.

Remark 3.6. $\mathcal{F}_{\mathfrak{c}} \subset \mathcal{F}_{\mathfrak{f}}$ since $G(x) \le 1/e$. When S and T are of comparable size, $\mathcal{F}_{\mathfrak{c}}$ requires $|S|, |T|$ to be of order at least $\sqrt{n}$, while $\mathcal{F}_{\mathfrak{f}}$ allows sizes down to order $\log n$; indeed, the condition in (3.13) rearranges to

$$\min(|S|, |T|) \ge K \log \frac{n}{\max(|S|, |T|)}.$$

We do not make use of the fine scale family $\mathcal{F}_{\mathfrak{f}}$ in this work, but include (3.16) for the sake of completeness. The bound (3.15) will be used in the proof of Proposition 2.6.

Remark 3.7. Theorem 3.5 goes beyond what can be obtained by the restriction method. Indeed, for an $n \times n$ 0/1 matrix B with i.i.d. $\mathrm{Ber}(\delta)$ entries, we have by Bernstein's inequality that

$$\mathbf{P}\left(\left| \frac{e_{S,T}(B)}{\delta |S||T|} - 1 \right| \ge \lambda \right) \le 2 \exp(-c\lambda^2 \delta^2 |S||T|). \tag{3.17}$$

From (1.4), this is only smaller than the probability $p_{\delta,n}$ that $B \in \mathcal{M}_d$ for

$$|S||T| \gg_\delta n \log n. \tag{3.18}$$

Remark 3.8. See Lemma 29 in [17] for a similar edge discrepancy property for sparse random regular graphs. The statement there is for undirected graphs, but the proof applies equally well to digraphs.

We note that an argument similar to the proof of Theorem 3.5 can be used to bound the largest nontrivial singular value $\sigma_2(M)$, which is stronger than the discrepancy property. However, we defer this result to a later work as it is too far from our current needs. The proofs of Proposition 3.1 and Theorem 3.5 are deferred to the appendix. The arguments there make use of Chatterjee's method of exchangeable pairs for concentration of measure [14]. The construction of couplings of r.r.d. matrices $(M, \widetilde{M})$, which will be used in Sections 4 and 5 as well as in the appendix, is the subject of the next section.

3.2. **Local couplings.** In this section we establish a number of ways of creating coupled pairs $(M, \widetilde{M})$ of matrices uniformly distributed in $\mathcal{M}_d$ by applying transformations to small minors of M. This general technique is known as the method of *switchings* and has been used extensively by McKay, Wormald and coauthors in several works on random regular graphs – see for instance section 2.4 of the survey [57]. The reader may also like to see [58] for a simple example of how the method can be used to control the event that a random permutation has a fixed point.

The couplings will be constructed from two basic transformations of M, which we call *switchings* and *reflections*. The couplings $(M, \widetilde{M})$ defined below will be invoked at several stages of the proof (including the appendix), where in order to estimate the probability that some event holds for M, we will replace M with $\widetilde{M}$ and proceed using the randomness of the switchings or reflections.

Definition 3.9 (Switching)**.** *We call a* 2×2 *minor* $M_{(i_1,i_2)\times(j_1,j_2)}$ switchable *if it is equal to either*

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

By perform a switching at $(i_1, i_2) \times (j_2, j_2)$ *on* M *we mean to replace the minor* $M_{(i_1,i_2)\times(j_1,j_2)}$ *with* $\mathbf{J}_2$ *if it is* $\mathbf{I}_2$*, and* $\mathbf{I}_2$ *if it is* $\mathbf{J}_2$*, and to leave* M *unchanged if this minor is not switchable.*

In the associated digraph G this corresponds to changing between the following edge configurations:

where we use solid arrows to depict directed edges, and dashed arrows to indicate places where there is no edge (i.e. "non-edges").

Definition 3.10 (Random switching)**.** *For ξ a $\mathrm{Ber}_\pm(1/2)$ random variable independent of M, by* perform a random switching at $(i_1,i_2)\times(j_2,j_2)$ according to ξ *on M, we mean to replace this minor with $\mathbf{I}_2$ if the minor is switchable and $\xi=+1$, $\mathbf{J}_2$ if the minor is switchable and $\xi=-1$, and to leave M unchanged if the minor is not switchable.*

One can think of a random switching as resampling the minor $M_{(i_1,i_2)\times(j_1,j_2)}$ uniformly at random conditional on the other n^2-4 entries of the matrix, with ξ encoding the outcome of the resampling when there are two alternatives for this minor.

Let
$$\Phi_{(i_1,i_2)\times(j_1,j_2)}:\mathcal{M}_d\to\mathcal{M}_d$$
denote the map which performs a switching at $(i_1,i_2)\times(j_2,j_2)$, and for ξ a $\mathrm{Ber}_\pm(1/2)$ random variable, let
$$\Phi^{\xi}_{(i_1,i_2)\times(j_1,j_2)}:\mathcal{M}_d\to\mathcal{M}_d$$
denote the map which performs a random switching at $(i_1,i_2)\times(j_2,j_2)$ according to ξ. Note that $\Phi_{(i_1,i_2)\times(j_1,j_2)}$ is an involution on $\mathcal{M}_d$.

Lemma 3.11 (Single switching)**.** *Let M be a uniform random element of $\mathcal{M}_d$, let I_1,I_2,J_1,J_2 be random elements of $[n]$ independent of M, and let ξ be a $\mathrm{Ber}_\pm(1/2)$ random variable independent of all other variables. Then*
$$\widetilde{M}_1:=\Phi_{(I_1,I_2)\times(J_1,J_2)}(M)$$
and
$$\widetilde{M}_2:=\Phi^{\xi}_{(I_1,I_2)\times(J_1,J_2)}(M)$$
are also uniformly distributed on $\mathcal{M}_d$. Moreoever, $(M,\widetilde{M}_1)$ is an exchangeable pair of random matrices.

Remark 3.12. Note that we have made no assumption on the distribution of the random indices I_1,I_2,J_1,J_2. In particular they may be deterministic.

Proof. Conditional on I_1,I_2,J_1,J_2, it follows from the fact that $\Phi_{(I_1,I_2)\times(J_1,J_2)}$ is an involution that $(M,\widetilde{M}_1)$ is an exchangeable pair, and in particular that $\widetilde{M}_1$ is uniformly distributed on $\mathcal{M}_d$.

As for $\widetilde{M}_2$, let us fix $M_0\in\mathcal{M}_d$. Our goal is to show that $\mathbf{P}(\widetilde{M}_2=M_0)=\mathbf{P}(M=M_0)$. Let $\mathcal{E}$ denote the event, in the randomness of I_1,I_2,J_1,J_2, that M_0 is switchable at $(I_1,I_2)\times(J_1,J_2)$. On $\mathcal{E}^c$ we have $\widetilde{M}_2=M$.

Now condition on I_1, I_2, J_1, J_2 such that $\mathcal{E}$ holds. Let $M_0^{\mathbf{I}_2}$ and $M_0^{\mathbf{J}_2}$ denote the matrices obtained from M_0 by replacing the $(I_1, I_2) \times (J_1, J_2)$ minor with $\mathbf{I}_2$ and $\mathbf{J}_2$, respectively. M_0 is equal to exactly one of these. Without loss of generality, suppose $M_0 = M_0^{\mathbf{I}_2}$. Now we have

$$\begin{aligned}\mathbf{P}\left(\widetilde{M}_2 = M_0 \middle| I_1, I_1, J_1, J_2\right)1_{\mathcal{E}} &= \mathbf{P}\left(\left\{M = M_0^{\mathbf{I}_2} \text{ or } M_0^{\mathbf{J}_2}\right\} \wedge \{\xi = +1\} \middle| I_1, I_2, J_1, J_2\right)1_{\mathcal{E}} \\ &= \frac{1}{2}\left(\mathbf{P}_M\left(M = M_0^{\mathbf{I}_2}\right) + \mathbf{P}_M\left(M = M_0^{\mathbf{J}_2}\right)\right)1_{\mathcal{E}} \\ &= \mathbf{P}(M = M_0)1_{\mathcal{E}}\end{aligned}$$

where in the last line we have used the fact that M is uniformly distributed.

Summarizing, we have

$$\begin{aligned}\mathbf{P}\left(\widetilde{M}_2 = M_0 \middle| I_1, I_2, J_1, J_2\right) &= \mathbf{P}\left(\widetilde{M}_2 = M_0 \middle| I_1, I_2, J_1, J_2\right)1_{\mathcal{E}} \\ &\quad + \mathbf{P}\left(\widetilde{M}_2 = M_0 \middle| I_1, I_2, J_1, J_2\right)1_{\mathcal{E}^c} \\ &= \mathbf{P}(M = M_0 | I_1, I_2, J_1, J_2)\end{aligned}$$

and undoing the conditioning on I_1, I_2, J_1, J_2 concludes the proof. □

We can iterate Lemma 3.11 to create couplings from several random switchings.

Corollary 3.13 (Several independent switchings)**.** *Let M be a uniform random element of $\mathcal{M}_d$ and $m \geq 1$. We randomly sample 2×2 minors*

$$\{M_{\mathbf{p}_l \times \mathbf{q}_l}\}_{l=1}^m$$

with $\{\mathbf{p}_l\}_{l=1}^m, \{\mathbf{q}_l\}_{l=1}^m$ chosen independently of M, and so that no entry of M appears in more than one minor $M_{\mathbf{p}_l \times \mathbf{q}_l}$ (note that this condition does not depend on M).

Let $\{\xi_l\}_{l=1}^m$ be i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables. Conditional on M, form $\widetilde{M}$ by performing a random switching according to ξ_l at each $\mathbf{p}_l \times \mathbf{q}_l$. Then $\widetilde{M}$ is uniformly distributed on $\mathcal{M}_d$.

Remark 3.14. Here again, $\mathbf{p}_l, \mathbf{q}_l$ can be deterministic as we have made no assumption on their distributions. Indeed, we will often apply this corollary to sample 2×2 minors from a fixed pair of columns (j_1, j_2), in which case we will take $\mathbf{q}_l = (j_1, j_2)$ for each $l \in [m]$.

Remark 3.15. One can similarly create uniformly distributed matrices $\widetilde{M}$ coupled to M without the restriction that each entry of M is sampled at most once, but in applications of Corollary 3.13 we want to avoid having to consider events on which an entry of M is edited multiple times.

Proof. Set $M^{(0)} = M$, and for each $m' \in [m]$ form $M^{(m')}$ by performing random switchings according to ξ_l at each pair $\mathbf{p}_l \times \mathbf{q}_l$ for $1 \leq l \leq m'$. By conditioning on $\{\mathbf{p}_l, \mathbf{q}_l\}_{l=1}^{m'-1}$ and $\{\xi_l\}_{l=1}^{m'-1}$ we see that $M^{(m')} \stackrel{d}{=} M^{(m'-1)}$ by Lemma 3.11. It inductively follows that

$$\widetilde{M} = M^{(m)} \stackrel{d}{=} M^{(0)} = M.$$

□

To prove the codegree discrepancy property of Proposition 3.1, we will need an additional "reflection" coupling whose definition is a little more subtle. This coupling is only applied in the appendix, so the reader may desire to skip to the next section on a first reading, taking the discrepancy properties as black boxes for the proof of Theorem 1.2.

Recall from the discussion in Section 2 that we will need the codegree discrepancy property of Proposition 3.1 in order to apply the switchings constructions above. Indeed, in order to apply several independent switchings on the first two rows using Corollary 3.13, we would like the sets

$$T^{10} = \left\{ j \in [n] : M_{(1,2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \quad T^{01} = \left\{ j \in [n] : M_{(1,2)\times j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

to be large, i.e. of size $\Omega(n)$. If this is the case, we can then locate switchable 2×2 minors by randomly sampling pairs of column indices (J_l^+, J_l^-). Conditional on the event that one of these lands in T^{10} and the other lands in T^{01}, the sampled minor $M_{(1,2)\times(J_l^+, J_l^-)}$ is switchable. According to Proposition 3.1, with overwhelming probability we have

$$|T^{10}| = (1 + O(\lambda))\delta(1-\delta)n$$

for any fixed $\lambda > 0$. From the row sums constraint we have that $|T^{10}| = |T^{01}|$, so the probability of sampling a switchable minor is $\Omega_\delta(1)$ as desired.

Now in order to prove Proposition 3.1, we will first prove a crude lower bound of the form

$$|T^{10}| \gg_\delta n \tag{3.19}$$

– see Lemma A.1. (3.19) is actually sufficient for the location of switchable minors by random sampling, but we will need the finer information provided by Proposition 3.1 in order to prove the edge discrepancy property of Theorem 3.5.

We now consider possible approaches to proving (3.19) in order to motivate the definition of reflections below. A naïve attempt might proceed by using Corollary 3.13 to apply several independent random switchings on the entries of the first two rows, and show that the resulting cardinality $|T^{10}|$ is binomially distributed. One would then deduce that these sets are large from a bound on the lower tail of the binomial distribution.

However, one immediately observes the $|T^{10}|$ is *invariant* under switchings on minors of $M_{(1,2)\times[n]}$, so the switchings would have to be between the second and third rows, say. Now considering switchings on $M_{(2,3)\times[n]}$, we are stuck with the problem that there may not be many switchable minors here either – i.e. that

$$\left\{ j \in [n] : M_{(2,3)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \quad \left\{ j \in [n] : M_{(2,3)\times j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

might be small. We hence see that an approach to getting a lower bound on $|T^{10}|$ using switchings is circular. The method of switchings cannot be employed until we know that the sets

$$T^{10}_{(i_1,i_2)} = \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

are large, so we require another coupling construction to prove this.

For this purpose we define the *reflection* coupling. The idea goes as follows. Note that if

$$T^{10} = \left\{ j \in [n] : M_{(1,2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

is small, it follows that most minors $M_{(1,2)\times j}$ are of the form $\binom{1}{1}$ or $\binom{0}{0}$, i.e. that

$$T^{11} = \left\{ j \in [n] : M_{(1,2)\times j} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad \text{and} \quad T^{00} = \left\{ j \in [n] : M_{(1,2)\times j} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

are large. We will show that this sort of imbalance is unlikely by showing that we can perform alterations between 2×2 minors of $M_{(1,2)\times[n]}$ of the form

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \longleftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.20}$$

The "forward" direction of such an alteration removes an element from T^{11} and adds it to T^{10}, while the "reverse" direction does the opposite. If we can perform several such alterations independently at non-overlapping minors $(1,2) \times (j_1, j_2)$, as in Corollary 3.13, then $|T^{10}|$ will be binomially distributed and we can easily conclude a lower bound that holds with overwhelming probability.

It remains to see how to construct a "legal" alteration on a minor $(1,2) \times (j_1, j_2)$ of M that has the result of (3.20), where by "legal" we mean that the resulting matrix $\widetilde{M}$ is still uniformly distributed in $\mathcal{M}_d$. Indeed, in order that the column sums constraint is still satisfied, such an alteration will have to be compensated by changes in other entries of columns j_1, j_2 of M.

Again it is instructive to first consider a naïve approach. Let us fix the columns (j_1, j_2) and suppose that we have

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

In order to accomplish the forward alteration of (3.20), we could look for the first row index $I \geq 3$ such that $M_{I\times(j_1,j_2)} = (0 \quad 1)$, and perform a switching at the minor $(2, I) \times (j_1, j_2)$ to send

$$M_{(1,2,I)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.21}$$

Note that such an index I must exist by the column sums constraint. We need to perform the reverse alteration as well, as otherwise we would be biasing the distribution of $\widetilde{M}$ to have T^{10} large and T^{11} small. For (j_1, j_2) such that

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \tag{3.22}$$

let $I(j_1, j_2)$ denote the first row index in $[3, n]$ such that $M_{I(j_1,j_2)\times(j_1,j_2)} = (0 \quad 1)$, and for (j_1, j_2) such that

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.23}$$

let $I(j_1, j_2)$ denote the first row index in $[3, n]$ such that $M_{I(j_1,j_2)\times(j_1,j_2)} = (1 \quad 0)$. Note that in this latter case, it may be that no such index exists – this is the event that $M_{i\times(j_1,j_2)}$ is either $(1 \quad 1)$ or $(0 \quad 0)$ for every $i \in [3, n]$. Set $I(j_1, j_2) = 2$ in this case.

Now let

$$\Psi_{(j_1,j_2)} : \mathcal{M}_d \to \mathcal{M}_d$$

denote the mapping that performs a switching at the minor $(2, I(j_2, j_2)) \times (j_1, j_2)$ if either (3.22) or (3.23) holds. Note that if $I(j_1, j_2) = 2$, this minor is not switchable and so $\Psi_{(j_1,j_2)}$ acts as the identity in this case.

It turns out that $\Psi_{(j_1,j_2)}$ defined in this way has some bias. Indeed, consider the following example configurations for $M_{[5]\times(j_1,j_2)}$:

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \hline 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \hline 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.24}$$

Note that while $\Psi_{(j_1,j_2)}$ sends A_0 to A_1, it sends all of A_1, A_2, A_3 to A_0. We may similarly define A_4, A_5, etc. which are sent to A_0.

$$\begin{tikzcd} A_1 \arrow[d, leftrightarrow] & A_2 \arrow[dl] & A_3 \arrow[dll] & \cdots \\ A_0 & & & \end{tikzcd} \tag{3.25}$$

The above diagram illustrates the fact that $\Psi_{(j_1,j_2)}$ is not a bijection on $\mathcal{M}_d$, so that $\Psi_{(j_1,j_2)}(M)$ is not uniformly distributed when M is uniformly distributed. We now modify our approach to the selection of $I(j_1,j_2)$ to define a bijection $\Xi_{(j_1,j_2)}$ (in fact it will be an *involution*).

Essentially, rather than perform a switching at $(2,I)\times(j_1,j_2)$, the idea is to interchange the columns of the *entire minor* $[2,I]\times(j_1,j_2)$, for an appropriate value of I. We want a rule for the selection of I such that the same value of I is selected for the transformed matrix $\Xi_{(j_1,j_2)}(M)$, giving

$$\Xi_{(j_1,j_2)}(\Xi_{(j_1,j_2)}(M)) = M.$$

The key is to associate the $n\times 2$ minor $M_{[n]\times(j_1,j_2)}$ with a *walk* on $\mathbf{Z}$. Starting at the origin, we look at each minor $M_{i\times(j_1,j_2)}$ in turn for $i=1,\dots,n$: if $M_{i\times(j_1,j_2)} = (1 \quad 0)$ we take a step to the left, if $M_{i\times(j_1,j_2)} = (0 \quad 1)$ we step to the right, and otherwise we do not move. To $M_{[n]\times(j_1,j_2)}$ we associate a vector $w\in\mathbf{Z}^n$, whose ith component records the position of the walk after i steps:

$$w(I) = \sum_{i=1}^{I} 1_{\left\{M_{i\times(j_1,j_2)}=(0 \quad 1)\right\}} - 1_{\left\{M_{i\times(j_1,j_2)}=(1 \quad 0)\right\}}. \tag{3.26}$$

By the column sums constraint, this walk ends at the origin, i.e. $w(n)=0$, and takes anywhere from 0 to $2d$ steps.

Definition 3.16 (Reflecting pair)**.** *With w as in (3.26), we say a pair of column indices* (j_1,j_2) reflecting *if*

(1) $w(1)=-1$,
(2) $w(2)\neq -1$, and
(3) there is some $i\geq 3$ such that $w(i)=-1$.

Let $I^*_{(j_1,j_2)} = I^*_{(j_1,j_2)}(M)$ be the smallest $i\geq 3$ such that $w(i)=-1$. In terms of the entries of M, conditions (1) and (2) mean that $M_{(1,2)\times(j_1,j_2)}$ either has the form (3.22) or (3.23),

and (3) means that there is some $I \geq 3$ such that

$$\sum_{i=2}^{I} M(i, j_1) = \sum_{i=2}^{I} M(i, j_2) \tag{3.27}$$

with $I^*_{(j_1,j_2)}$ denoting the smallest such I.

We note that if (3.22) holds, then $w(2) = -2$ and so condition (3) holds automatically as the walk must pass through -1 on its way back to the origin. However, if (3.23) holds we have $w(2) = 0$ and condition (3) need not hold.

Definition 3.17 (Reflection)**.** *For $S \subset [n]$ and $(j_1, j_2) \in [n] \times [n]$, by* perform a reflection at (j_1, j_2) on M *we mean to replace the minor $M_{[2,I^*(j_1,j_2)]\times(j_1,j_2)}$ with the "reflected" minor $M_{[2,I^*(j_1,j_2)]\times(j_2,j_1)}$ (with columns interchanged) if (j_1, j_2) is reflecting, and to leave M unchanged if this pair is not reflecting.*

For ξ a $\mathrm{Ber}_\pm(1/2)$ *random variable independent of M, by* perform a random reflection at (j_1, j_2) according to ξ on M *we mean to replace the minor $M_{[2,I^*(j_1,j_2)]\times(j_1,j_2)}$ with the reflected minor $M_{[2,I^*(j_1,j_2)]\times(j_2,j_1)}$ if (j_1, j_2) is reflecting and $\xi = +1$, and to do nothing otherwise.*

Note that while for the random switching $\Phi^{\xi}_{(i_1,i_2)\times(j_1,j_2)}$, ξ encoded the *outcome* of the switching, here ξ encodes whether or not a reflection is performed. The distinction is of no major importance and has been made only for notational convenience.

Denote by

$$\Xi_{(j_1,j_2)} : \mathcal{M}_d \to \mathcal{M}_d$$

the map that performs a reflection at (j_1, j_2), and for ξ a $\mathrm{Ber}_\pm(1/2)$ random variable let

$$\Xi^{\xi}_{(j_1,j_2)} = 1_{\{\xi=+1\}}\Xi_{(j_1,j_2)} + 1_{\{\xi=-1\}}\,\mathrm{Id} \tag{3.28}$$

denote the random map that performs a random reflection at $(j_1, j_2)]$ according to ξ.

For a reflecting pair (j_1, j_2), the mapping $\Xi_{(j_1,j_2)}$ has the desired effect on the first two rows of M, as we have

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad\Longrightarrow\quad \Xi_{(j_1,j_2)}(M)_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.29}$$

and

$$M_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad\Longrightarrow\quad \Xi_{(j_1,j_2)}(M)_{(1,2)\times(j_1,j_2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}. \tag{3.30}$$

Now we argue that $\Xi_{(j_1,j_2)}$ is an involution on $\mathcal{M}_d$. Consider the effect of performing a reflection at (j_1, j_2) on the associated walk. Let $w \in \mathbf{Z}^n$ be the walk associated to $M_{[n]\times(j_1,j_2)}$, and $\tilde{w}$ be the walk associated to $\Xi_{(j_1,j_2)}(M)_{[n]\times(j_1,j_2)}$. We have

$$w(1) = w(I^*_{(j_1,j_2)}(M)) = -1$$

and by definition of $I^*_{(j_1,j_2)}$,

$$w(i) \neq -1 \quad \text{for all } 2 < i < I^*_{(j_1,j_2)}(M).$$

Now the reflection changes the direction of all steps in $[2, I^*_{(j_1,j_2)}(M)]$, reflecting this part of the walk across -1. Hence we have

$$\tilde{w}(1) = \tilde{w}(I^*_{(j_1,j_2)}(M)) = -1$$

and

$$\tilde{w}(i) \neq -1 \quad \text{for all } 2 < i < I^*_{(j_1,j_2)}(M).$$

As a consequence, $I^*_{(j_1,j_2)}(M)$ is the first $i \geq 3$ such that $\tilde{w}(i) = -1$, and so we have

$$I^*_{(j_1,j_2)}(\Xi_{(j_1,j_2)}(M)) = I^*_{(j_1,j_2)}(M). \tag{3.31}$$

It follows from the definition of $\Xi_{(j_1,j_2)}$ that it is an involution. By (3.28) we have that $\Xi^{\xi}_{(j_1,j_2)}$ is an involution conditional on ξ. As an immediate consequence we have

Lemma 3.18 (Reflection coupling)**.** *Let M be a uniform random element of $\mathcal{M}_d$, let (J_1, J_2) be sampled randomly from $[n] \times [n]$ independently of M, and let ξ be a* $\mathrm{Ber}_{\pm}(1/2)$ *random variable independent of M and (J_1, J_2). Then*

$$\widetilde{M}_1 := \Xi_{(J_1,J_2)}(M)$$

and

$$\widetilde{M}_2 := \Xi^{\xi}_{(J_1,J_2)}(M)$$

are uniformly distributed on $\mathcal{M}_d$.

By arguing as in the proof of Corollary 3.13, we can iterate the above construction to get a coupled matrix $\widetilde{M}$ with more randomness injected, which we will need for better tail estimates.

Corollary 3.19 (Several independent reflections)**.** *Let M be a uniform random element of $\mathcal{M}_d$ and $m \geq 1$. Sample $\{J_k\}_{k=1}^{2m}$ without replacement from $[n]$ and independently of M. For each $l \in [m]$ set*

$$\mathbf{q}_l = (J_{2l-1}, J_{2l})$$

and let $\xi_1, \ldots, \xi_m$ be i.i.d. $\mathrm{Ber}_{\pm}(1/2)$ *random variables independent of all other variables. Form $\widetilde{M}$ by performing a random reflection at $\mathbf{q}_l$ according to ξ_l for each $l \in [m]$. Then $\widetilde{M}$ is uniformly distributed on $\mathcal{M}_d$.*

3.3. **Sampling pairs of row and column indices.** Having established all of the switchings constructions that we will use, we collect some simple concentration estimates that will allow us to locate a lot of switchable 2×2 minors by random sampling of pairs of row and column indices. We have the following well-known concentration estimate is due to Maurey [30]:

Theorem 3.20 (Concentration for the symmetric group)**.** *Let $N \geq 1$, let σ be a uniform random element of* $\mathrm{Sym}(N)$*, and let $F : \mathrm{Sym}(N) \to \mathbf{R}$ be a function that is 1-Lipschitz in the transposition distance, i.e.*

$$|F(\pi) - F(\tau \circ \pi)| \leq 1$$

for any $\pi \in \mathrm{Sym}(N)$ and any transposition $\tau \in \mathrm{Sym}(N)$. Then for any $t \geq 0$,

$$\mathbf{P}(F(\sigma) \geq \mathbf{E}\, F(\sigma) + t) \leq \exp\left(-ct^2/N\right) \tag{3.32}$$

for some absolute constant $c > 0$.

Proof. See also [27], p. 70, where this is deduced from a concentration inequality later proved by Schechtman [42] for a general class of finite metric spaces. Either way the proof comes down to an application of Azuma's inequality. $\square$

From the above theorem we can deduce a tail bound for Lipschitz functions of sequences of natural numbers sampled without replacement.

Corollary 3.21 (Concentration for sampling without replacement)**.** *Let $Y = (Y_1, \dots, Y_m)$ have components that are sequentially sampled uniformly without replacement from $[N]$, and let $F : [N]^m \to \mathbf{R}$ be a 1-Hamming Lipschitz function. Then for any $t \geq 0$,*

$$\mathbf{P}(F(Y) \geq \mathbf{E}\, F(Y) + t) \leq \exp\left(-ct^2/m\right) \tag{3.33}$$

for some absolute constant $c > 0$.

Proof. Let $\sigma \in \mathrm{Sym}(N)$ be a uniform random permutation, and observe that Y is identically distributed to $Z := (\sigma(1), \dots, \sigma(m))$. Now $F(Z)$ is a 2-Lipschitz function on $\mathrm{Sym}(N)$ with the respect to the transposition distance. We may now apply Theorem 3.20 (after scaling F by $1/2$ to get a 1-Lipchitz function). □

In the proof of Theorem 1.2 we will sample pairs of indices in two different ways. The following corollary will allow us to argue that with either method of sampling, the number of "good" sampled pairs is large, and the number of "bad" sampled pairs is small, where "good" and "bad" will depend on the context.

Corollary 3.22 (Bounds on good and bad samples)**.** *Let $n, m \geq 1$, and consider the following two ways of sampling m ordered pairs of indices from $[n]$:*

(1) Sample $I_1, \dots, I_{2m}$ uniformly without replacement from $[n]$, and for each $l \in [m]$ put
$$\mathbf{p}_l = (I_{2l-1}, I_{2l}).$$
(2) Let $k \in [m, n-m]$, sample $J_1^+, \dots, J_m^+$ uniformly without replacement from $[k]$, and independently of these sample $J_1^-, \dots, J_m^-$ uniformly without replacement from $[k+1, n]$. For each $l \in [m]$ put
$$\mathbf{q}_l = (J_l^+, J_l^-).$$

Let $E \subset [n] \times [n]$, and set

$$Z_{\mathcal{P}} = |\{l : \mathbf{p}_l \in E\}|, \quad Z_{\mathcal{Q}} = |\{l : \mathbf{q}_l \in E\}|.$$

Then

(1) If $\mathbf{P}(\mathbf{p}_l \in E) \leq p_1 \ \forall l \in [m]$, then $Z_{\mathcal{P}} \ll p_1 m$ with probability $1 - O(\exp(-cp_1^2 m))$.

(2) If $\mathbf{P}(\mathbf{p}_l \in E) \geq p_2 \ \forall l \in [m]$, then $Z_{\mathcal{P}} \gg p_2 m$ with probability $1 - O(\exp(-cp_2^2 m))$.

(3) If $\mathbf{P}(\mathbf{q}_l \in E) \leq q_1 \ \forall l \in [m]$, then $Z_{\mathcal{Q}} \ll q_1 m$ with probability $1 - O(\exp(-cq_1^2 m))$.

(4) If $\mathbf{P}(\mathbf{q}_l \in E) \geq q_2 \ \forall l \in [m]$, then $Z_{\mathcal{Q}} \gg q_2 m$ with probability $1 - O(\exp(-cq_2^2 m))$.

Proof. For (1), by linearity of expectation we have $\mathbf{E}\, Z_{\mathcal{P}} \leq p_1 m$. Now note that $Z_{\mathcal{P}}$ is a 1-Hamming Lipschitz function of $(I_1, \dots, I_{2m})$. From Corollary 3.21 we have

$$\begin{aligned}\mathbf{P}(Z_{\mathcal{P}} \geq (1+\varepsilon)p_1 m) &\leq \mathbf{P}(Z_{\mathcal{P}} \geq \mathbf{E}\, Z_{\mathcal{P}} + \varepsilon p_1 m) \\ &\leq \exp\left(-c\varepsilon^2 p_1^2 m\right)\end{aligned}$$

which is acceptable for any fixed $\varepsilon > 0$. We can obtain the lower tail estimate (2) in a similar way by applying Corollary 3.21 to the 1-Lipschitz function $-Z_{\mathcal{P}}$.

For (3) and (4), let $\mathcal{Y}_+, \mathcal{Y}_-$ denote the probability spaces of sequences of length m sampled without replacement from $[k]$ and $[k+1,n]$, respectively, equipped with the uniform measures. Now we note that $Z_{\mathcal{Q}}$ is a 1-Hamming Lipschitz function on the product probability space $\mathcal{Y} = \mathcal{Y}_+ \otimes \mathcal{Y}_-$. Letting

$$J^+ = (J_1^+, \dots, J_m^+) \in \mathcal{Y}_+, \quad J^- = (J_1^-, \dots, J_m^-) \in \mathcal{Y}_-$$

be independent uniform random elements, we have

$$Z_{\mathcal{Q}} = f(J^+, J^-)$$

where $f(\cdot, j_-)$ and $f(j_+, \cdot)$ are 1-Lipschitz on $\mathcal{Y}_+$ and $\mathcal{Y}_-$, respectively, for each $j_+ \in [k], j_- \in [k+1,n]$. It follows that $g(j_-) = \mathbf{E}_{J^+} f(J^+, j_-)$ is a 1-Lipschitz function on $\mathcal{Y}_-$. Then for any $\varepsilon \geq 0$, by pigeonholing and a union bound we have

$$\begin{aligned}
\mathbf{P}(Z_{\mathcal{Q}} \geq 2q_1 m) &\leq \mathbf{P}\big(f(J^+,J^-) - \mathbf{E} f(J^+,J^-) \geq q_1 m\big) \\
&= \mathbf{P}\big(f(J^+,J^-) - \mathbf{E}_{J^+} f(J^+,J^-) + \mathbf{E}_{J^+} f(J^+,J^-) - \mathbf{E} f(J^+,J^-) \geq q_1 m\big) \\
&\leq \mathbf{E}_{J^-} \mathbf{P}\bigg(f(J^+,J^-) - \mathbf{E}_{J^+} f(J^+,J^-) \geq \frac{1}{2} q_1 m \big| J^-\bigg) \\
&\qquad\qquad + \mathbf{P}\bigg(g(J^-) - \mathbf{E}\, g(J^-) \geq \frac{1}{2} q_1 m\bigg) \\
&\leq 2\exp\left(-c q_1^2 m\right)
\end{aligned}$$

where in the last line we have applied Corollary 3.21 to each term. We can similarly obtain the lower tail estimate (4) by replacing $Z_{\mathcal{Q}}$ with $-Z_{\mathcal{Q}}$ and arguing as above. $\square$

4. High level proof of Theorem 1.2

In this section we prove Theorem 1.2, taking as black boxes Proposition 2.7 ruling out structured null vectors, and the codegree discrepancy property of Proposition 3.1 (Theorem 3.5 for edge discrepancy is not needed until Section 5).

Define the events

$$\mathcal{R}_r = \{\mathrm{rank}(M) \leq n - r\}.$$

Our goal is to bound $\mathbf{P}(\mathcal{R}_1)$. We will argue separately that $\mathbf{P}(\mathcal{R}_2)$ and $\mathbf{P}(\mathcal{R}_1 \setminus \mathcal{R}_2)$ are small.

Recall from Definition 2.5 that a vector $x \in \mathbf{R}^n$ has the small level sets property with parameter $\varepsilon > 0$, or $\mathrm{SLS}(\varepsilon)$ for short, if for any $\lambda \in \mathbf{R}$ we have $|x^{-1}(\lambda)| \leq \varepsilon n$. Recall that for $S \subset [n]$, M_S denotes the matrix obtained by removing the rows indexed by S. Define the event

$$\mathcal{G}_\varepsilon := \bigwedge_{\{i_1,i_2\} \subset [n]} \left\{\text{all left and right null vectors of } M_{\{i_1,i_2\}} \text{ and } (M^\mathsf{T})_{\{i_1,i_2\}} \text{ have } \mathrm{SLS}(\varepsilon)\right\}. \tag{4.1}$$

On this event we also have that any null vector of M has $\mathrm{SLS}(\varepsilon)$, since any null vector of M is also a null vector of $M_{\{i_1,i_2\}}$. By Proposition 2.7 and a union bound, $\mathcal{G}_\varepsilon$ holds with overwhelming probability if we take

$$\varepsilon := K_\delta n^{-1/8} \tag{4.2}$$

for some $K_\delta > 0$ sufficiently large depending only on δ.

The following lemma allows us to control the event $\mathcal{R}_r$ by a more structured event. We will only use it for the cases $r = 1, 2$. This is an extension of the double counting argument that was used to deduce (2.2) in Komlós' proof from Section 2.

Lemma 4.1. *Let $\mathcal{R}_{[r]}$ denote the event that $\mathcal{G}_\varepsilon$ holds, and that the columns $X_1, \dots, X_r$ of M all lie in the span of the remaining columns, which we denote V_r. Then for $r\varepsilon < 1/2$ we have*

$$\mathbf{P}(\mathcal{R}_r \wedge \mathcal{G}_\varepsilon) \le \frac{1}{1 - 2r\varepsilon} \mathbf{P}\big(\mathcal{R}_{[r]}\big). \tag{4.3}$$

Proof. For $T \subset [n]$, let

$$\mathcal{R}_T = \mathcal{G}_\varepsilon \wedge \big\{ X_j \in \operatorname{span}(X_{j'} : j' \notin T) \ \ \forall j \in T \big\} .$$

Let $\{J_1, \dots, J_r\} \subset [n]$ be a uniform random subset of size r, drawn independently of M. By column exchangeability we have

$$\mathbf{P}\big(\mathcal{R}_{\{J_1,\dots,J_r\}}\big) = \mathbf{P}\big(\mathcal{R}_{[r]}\big).$$

Now since

$$\mathbf{P}\big(\mathcal{R}_{\{J_1,\dots,J_r\}}\big) = \mathbf{P}\big(\mathcal{R}_{\{J_1,\dots,J_r\}} | \mathcal{R}_r \wedge \mathcal{G}_\varepsilon\big) \mathbf{P}(\mathcal{R}_r \wedge \mathcal{G}_\varepsilon)$$

it suffices to show that conditional on M in $\mathcal{R}_r \wedge \mathcal{G}_\varepsilon$,

$$\mathbf{P}_{J_1,\dots,J_r}\big(\mathcal{R}_{\{J_1,\dots,J_r\}} | M\big) \ge 1 - 2r\varepsilon. \tag{4.4}$$

We view the tuple $(J_1, \dots, J_r)$ as being drawn sequentially: J_1 is drawn uniformly from $[n]$, and for each $2 \le k \le r$, conditional on $J_1, \dots, J_{k-1}$, J_k is drawn uniformly from $[n] \setminus \{J_1, \dots, J_{k-1}\}$.

On $\mathcal{R}_r \wedge \mathcal{G}_\varepsilon$, there is a subspace W of the nullspace of M of dimension r such that each nonzero $x \in W$ has the small level sets property with parameter ε. Fix a basis $u_1, \dots, u_r$ of W. Let U denote the $r \times n$ matrix with rows $u_1^\mathsf{T}, \dots, u_r^\mathsf{T}$.

Viewing the members u_i of this basis as giving coefficients for a linear dependency among the columns of M, our goal is to show that with probability at least $1 - 2r\varepsilon$, we can row reduce U to a matrix V of basis vectors such that the minor $V_{[r] \times (J_1,\dots,J_r)}$ is upper triangular with 1s on the diagonal. From this it follows that the columns $X_{J_1}, \dots, X_{J_r}$ can simultaneously be written as linear combinations of the remaining $n - r$ columns.

Set $v_1 = u_1$. Let $\mathcal{B}_1$ be the event that $v_1(J_1) = 0$. Since v_1 has SLS(ε), $\mathbf{P}(\mathcal{B}_1) \le \varepsilon$. On $\mathcal{B}_1^c$, we define $v_1' = \frac{1}{v_1(J_1)} v_1$.

For $2 \le k \le r$, having defined $v_1', \dots, v_{k-1}'$ and bad events $\mathcal{B}_1, \dots, \mathcal{B}_{k-1}$, we set $v_k = u_k - \sum_{i=1}^{k-1} u_k(J_i) v_i'$. Let

$$\mathcal{B}_k = \{v_k(J_k) = 0\} \wedge \bigwedge_{i=1}^{k-1} \mathcal{B}_i^c.$$

Since $v_k \in W$ has SLS(ε) we have

$$\mathbf{P}(\mathcal{B}_k | J_1, \dots, J_{k-1}) \le 2\varepsilon,$$

for $k \le n/2$.

On $(\bigvee_{k=1}^{r} \mathcal{B}_k)^c$, we end with a matrix V with rows $v_i'^{\mathsf{T}}$ with the desired property. Hence, the probability that $X_{J_1}, \ldots, X_{J_r}$ are all in the span of the remaining columns, where $(J_1, \ldots, J_r)$ is a uniform random ordered m-tuple of elements of $[n]$, is bounded by

$$\sum_{k=1}^{r} \mathbf{P}(\mathcal{B}_k) \leq 2r\varepsilon.$$

$\square$

First we bound $\mathbf{P}(\mathcal{R}_2)$. By Lemma 4.1,

$$\begin{aligned} \mathbf{P}(\mathcal{R}_2) &\leq \mathbf{P}(\mathcal{G}_\varepsilon^c) + \mathbf{P}(\mathcal{R}_2 \wedge \mathcal{G}_\varepsilon) \\ &\leq \mathbf{P}(\mathcal{G}_\varepsilon^c) + \frac{1}{1-4\varepsilon}\mathbf{P}(\mathcal{R}_{[2]}) \end{aligned} \tag{4.5}$$

where we recall that $\mathcal{R}_{[2]}$ is the event that $\mathcal{G}_\varepsilon$ holds and the first two columns X_1, X_2 of M lie in the span of the remaining columns $V_2 = \operatorname{span}(X_3, \ldots, X_n)$. With $\varepsilon = K_\delta n^{-1/8}$, it only remains to bound $\mathbf{P}(\mathcal{R}_{[2]})$.

Note that on $\mathcal{R}_{[2]}$, X_1 and X_2 are orthogonal to any vector normal to V_2. Choose $u \in V_2^\perp$ arbitrarily (say chosen uniformly at random from the unit sphere in $V_2^\perp$, conditional on $X_3, \ldots, X_n$). We may assume that u has $\mathrm{SLS}(\varepsilon)$. Indeed, the alternative is empty on $\mathcal{R}_{[2]}$, since on this event u is orthogonal to every column of M and is hence a left null vector, putting us in the complement of $\mathcal{G}_\varepsilon$.

By Proposition 3.1,

$$h_{(1,2)}(M^{\mathsf{T}}) \gg v_\delta n \tag{4.6}$$

with overwhelming probability, so we may restrict to the event that (4.6) holds.

We define a coupling $(M, \widetilde{M})$ of r.r.d. matrices as follows. Let $m = \eta n$ for some $\eta > 0$ to be chosen later, and sequentially sample $\{I_l\}_{l=1}^{2m}$ uniformly without replacement from n and independently of M. For each $l \in [m]$, let

$$\mathbf{p}_l = (I_l^+, I_l^-) := (I_{2l-1}, I_{2l})$$

and denote also

$$\mathcal{P} = (\mathbf{p}_l)_{l=1}^m.$$

Let $(\xi_l)_{l=1}^m$ be i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables independent of all other variables in play. Conditional on M and $\mathcal{P}$, form $\widetilde{M}$ by applying a random switching at $(I_l^+, I_l^-) \times (1,2)$ according to ξ_l for each $l \in [m]$. $\widetilde{M}$ is identically distributed to M by Corollary 3.13. Denote by $\widetilde{X}_1, \widetilde{X}_2$ the first two columns of $\widetilde{M}$.

For each $l \in [m]$, let

$$L = \{l \in [m] : M_{\mathbf{p}_l \times (1,2)} \text{ is switchable}\}.$$

Now we can write

$$\begin{pmatrix} \widetilde{X}_1 \cdot u \\ \widetilde{X}_2 \cdot u \end{pmatrix} = W_0 + \sum_{l \in L} a_l(u) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \xi_l \partial_l(u) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{4.7}$$

where W_0 collects all of the summands not coming from switchable minors, and we have set

$$a_l(u) = \frac{u(I_l^+) + u(I_l^-)}{2} \tag{4.8}$$

and

$$\partial_l(u) = \frac{u(I_l^+) - u(I_l^-)}{2}. \tag{4.9}$$

Let

$$L^* = \{l \in L : \partial_l(u) \neq 0\}.$$

If we can get a good lower bound on $|L^*|$, we can condition on M and $\mathcal{P}$ and apply Theorem 2.2 to get a good upper bound on $\mathbf{P}(\widetilde{X}_1 \cdot u = 0)$.

To that end, we first show that $|L|$ is large. By the lower bound (4.6) on $h_{(1,2)}(M^\mathsf{T})$, we have

$$\mathbf{P}\left(M_{\mathbf{p}_l \times (1,2)} \text{ is switchable}\big|(\mathbf{p}_{l'})_{l'=1}^{l-1}\right) \geq cv_\delta^2 - \eta$$

for each $l \in [m]$ and some $c > 0$. Taking η to be a sufficiently small multiple of v_δ^2, from Lemma 3.22 we have that

$$\mathbf{P}(|L| < c_1 v_\delta^4 n) \ll \exp(-cv_\delta^6 n) \tag{4.10}$$

for some absolute constant $c_1 > 0$.

Now letting

$$\mathcal{B}_l = \{u(I_l^+) = u(I_l^-)\}$$

we show that

$$Z = \sum_{l=1}^m 1_{\mathcal{B}_l}$$

is small. Since we are on $\mathcal{G}_\varepsilon$, we have

$$\mathbf{P}(\mathcal{B}_l) \leq \varepsilon$$

for each $l \in [m]$. By Lemma 3.22 it follows that there is an absolute constant $C_1 > 0$ such that

$$\mathbf{P}(Z > C_1 \varepsilon v_\delta^2 n) \ll \exp(-c\varepsilon^2 v_\delta^2 n) \ll \exp(-c_\delta n^{7/8}). \tag{4.11}$$

Now we have shown that

$$\mathcal{G}^* := \left\{h_{(1,2)}(M^\mathsf{T}) \geq cv_\delta n\right\} \wedge \left\{|L| \geq c_1 v_\delta^4 n\right\} \wedge \left\{Z \leq C_1 \varepsilon v_\delta^2 n\right\}. \tag{4.12}$$

holds with overwhelming probability, and that on $\mathcal{G}^*$ we have

$$\begin{aligned} |L^*| &\geq |L| - Z \\ &\geq (c_1 v_\delta^2 - C_1 \varepsilon) v_\delta^2 n \qquad (4.13) \\ &\gg_\delta n \qquad (4.14) \end{aligned}$$

since $\varepsilon = o(1)$.

We can write

$$\begin{aligned} \mathbf{P}(\mathcal{R}_{[2]}) &= \mathbf{P}(\mathcal{G}^{*c}) + \mathbf{P}\big(\mathcal{R}_{[2]} \wedge \mathcal{G}^*\big) \\ &= \mathbf{P}(\mathcal{G}^{*c}) + \mathbf{E}\,\mathbf{P}\Big(\{\widetilde{X}_1 \cdot u = \widetilde{X}_2 \cdot u = 0\} \wedge \mathcal{G}^* | X_3, \ldots, X_n\Big). \qquad (4.15) \end{aligned}$$

Conditioning on M and $\mathcal{P}$ such that $\mathcal{G}^*$ holds, the only randomness left is in the ξ_l variables. Considering the first component in (4.7), it follows from Theorem 2.2 and the bound (4.14) that

$$\mathbf{P}\Big(\widetilde{X}_1 \cdot u = 0 \big| M, \mathcal{P}\Big) 1_{\mathcal{G}^*} = O_\delta(n^{-1/2}).$$

Undoing the conditioning on $M, \mathcal{P}$, we conclude from (4.15) that

$$\mathbf{P}(\mathcal{R}_{[2]}) = O_\delta(n^{-1/2})$$

and hence by (4.5),

$$\mathbf{P}(\mathcal{R}_2) = O_\delta(n^{-1/2}). \tag{4.16}$$

It only remains to bound $\mathbf{P}(\mathcal{R}_1 \setminus \mathcal{R}_2)$. In particular we assume that M has at most one (right) null vector, up to dilation. Again by Lemma 4.1 it suffices to bound $\mathbf{P}(\mathcal{R}_{[1]} \setminus \mathcal{R}_2)$.

On $\mathcal{R}_{[1]}$, X_1 is orthogonal to any vector in $V_1^\perp$, where $V_1 = \mathrm{span}(X_2, \dots, X_n)$. We pick a random vector Y in $V_1^\perp$ as follows. Choose J uniformly at random from $[2, n]$, independent of M, and denote $V_{(1,J)} = \mathrm{span}(X_j : j \notin \{1, J\})$. Conditional on J, pick $u_1 \perp u_2$ unit vectors in $V_{(1,J)}^\perp$, say uniformly at random. Now let $Y = (X_J \cdot u_2)u_1 - (X_J \cdot u_1)u_2 \in V_1^\perp$. On $\mathcal{R}_{[1]}$ we have

$$\begin{aligned} 0 &= X_1 \cdot Y \\ &= (X_1 \cdot u_1)(X_J \cdot u_2) - (X_1 \cdot u_2)(X_J \cdot u_1). \end{aligned} \tag{4.17}$$

Conditional on J, we now construct a matrix $\widetilde{M}$ coupled to M using the randomly sampled pairs of rows $\mathcal{P}$ and Bernoulli variables $(\xi_l)_{l=1}^m$ as above, except we perform switchings at $\mathbf{p}_l \times (1, J)$ rather than $\mathbf{p}_l \times (1, 2)$. Let

$$L_J = \big\{ l \in [m] : M_{\mathbf{p}_l \times (1,J)} \text{ is switchable} \big\}$$

and let

$$S = \bigcup_{l \in L_J} \{I_l^+, I_l^-\}.$$

For u any fixed unit vector we can write

$$\begin{aligned} \begin{pmatrix} \widetilde{X}_1 \cdot u \\ \widetilde{X}_J \cdot u \end{pmatrix} = \sum_{i \in [n] \setminus S} \begin{pmatrix} X_1(i) \\ X_J(i) \end{pmatrix} u(i) + \sum_{i \in S} \frac{X_1(i) + X_J(i)}{2} u(i) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ + \sum_{l \in L_J} \xi_l \frac{u(I_l^+) - u(I_l^-)}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned} \tag{4.18}$$

Let

$$\begin{aligned} A_J(u) &= (X_1 + X_J) \cdot u, \\ \Delta_{J,S}(u) &= \frac{1}{2}(X_J - X_1)1_S \cdot u \end{aligned}$$

and $\partial_l(u)$ as in (4.9). We can re-express (4.18) as

$$\begin{pmatrix} \widetilde{X}_1 \cdot u \\ \widetilde{X}_J \cdot u \end{pmatrix} = \begin{pmatrix} X_1 \cdot u + \Delta_{J,S}(u) + \sum_{l \in L_J} \xi_l \partial_l(u) \\ X_J \cdot u - \Delta_{J,S}(u) - \sum_{l \in L_J} \xi_l \partial_l(u) \end{pmatrix}.$$

This combines with (4.17) to give (after some algebra)

$$\begin{aligned} \widetilde{X}_1 \cdot \widetilde{Y} = X_1 \cdot Y + (\Delta_{J,S}(u_1) A_J(u_2) - \Delta_{J,S}(u_2) A_J(u_1)) \\ + \sum_{l \in L_J} \xi_l (A_J(u_2) \partial_l(u_1) - A_J(u_1) \partial_l(u_2)) \\ =: s_0(J, S) + \sum_{l \in L_J} \xi_l s_l \end{aligned} \tag{4.19}$$

where we note that the terms summarized by $s_0(J,S)$ and s_l are deterministic under conditioning on J and $\mathcal{P}$. To prove the theorem it suffices to get a suitable lower bound on the number of nonzero steps s_l and apply Theorem 2.2.

First we use the randomness of J to deal with the event

$$\begin{aligned}\mathcal{B}_0 &= \{A_J(u_1) = A_J(u_2) = 0\}\\ &= \{X_1 + X_J \perp u_1 \text{ and } u_2\}\end{aligned}$$

on which *all* of the steps s_l are zero. Since $u_1, u_2 \in V_{(1,J)}^{\perp}$, this implies that they are left null vectors of the $n \times n-1$ matrix

$$M' = (X_1 + X_J, X_2, \dots, X_{J-1}, X_{J+1}, \dots, X_n)$$

obtained from M by deleting the Jth column and adding it to the first column. It follows that M' has a nontrivial right null vector $y = (y(1), y')$, which in turn implies that M has a right null vector $\tilde{y}$ whose restriction to the indices $[n] \setminus \{1, J\}$ is y', and whose 1st and Jth components are $y(1)$. We have hence shown that the event $\mathcal{B}_0$ is contained in the event that M has a right null vector with 1st and Jth components equal to each other – call this latter event $\mathcal{B}_0'$.

On $\mathcal{R}_{[1]} \setminus \mathcal{R}_2$, M has exactly one right null vector $\tilde{y}$, so we may condition on it. Now since we are on $\mathcal{G}_\varepsilon$ (included in the definition of $\mathcal{R}_{[1]}$), $\tilde{y}$ has SLS(ε). It follows that

$$\begin{aligned}\mathbf{P}(\mathcal{B}_0') &\le \mathbf{P}_J(\tilde{y}(1) = \tilde{y}(J))\\ &\le \varepsilon\\ &= K_\delta n^{-1/8}.\end{aligned} \tag{4.20}$$

Now we condition on some J outside the event $\mathcal{B}_0'$. Let

$$v = A_J(u_2)u_1 - A_J(u_1)u_2 \in V_{(1,J)}^{\perp}.$$

v is nontrivial since we are outside of the event $\mathcal{B}_0'$. As v is a left null vector of the matrix M'' obtained by deleting columns 1 and J from M, it follows from our restriction to $\mathcal{G}_\varepsilon$ that v has SLS(ε).

Now note that $s_l = \partial_l(v)$. We may argue exactly as was done above to bound $\mathbf{P}(\mathcal{R}_2)$, defining Z and $\mathcal{G}^*$ as in (4.11), (4.12) to conclude that

$$|\{l \in L_J : \partial_l(v) \neq 0\}| \gg_\delta n$$

with overwhelming probability. Applying Theorem 2.2 to the random walk (4.19), it follows that

$$\begin{aligned}\mathbf{P}\big(\mathcal{R}_{[1]} \setminus \mathcal{R}_2\big) &\le \mathbf{P}(\mathcal{B}_0') + \mathbf{P}\left(\left\{\sum_{l \in L_J} \xi_l s_l = 0\right\} \setminus \mathcal{B}_0'\right)\\ &\le K_\delta n^{-1/8} + O_\delta(n^{-1/2})\\ &= O_\delta(n^{-1/8}).\end{aligned}$$

By Lemma 4.1 we have

$$\mathbf{P}(\mathcal{R}_1 \setminus \mathcal{R}_2) = O_\delta(n^{-1/8})$$

and combining this with (4.16) completes the proof. $\square$

5. Proof of Proposition 2.7 (No structured null vectors)

We recall our terminology from Section 2. We say that a vector $x \in \mathbf{R}^n$ has the small level sets property with parameter ε, abbreviated SLS(ε), if for any $\lambda \in \mathbf{R}$ we have that $|x^{-1}(\lambda)| \leq \varepsilon n$, where

$$x^{-1}(\lambda) := \{i \in [n] : x(i) = \lambda\}.$$

Our aim in this section is to prove the following:

Proposition 2.7 (No structured null vectors for M_S)**.** *For $S \subset [n]$, let M_S denote an $(n - |S|) \times n$ matrix obtained from M by deleting the rows indexed by S. For any $r \geq 0$ there is a constant $K_{\delta,r} > 0$ depending only on δ, r such that for any fixed $S \subset [n]$ with $|S| = r$, with overwhelming probability any nontrivial right null vector of M_S has* SLS$(K_{\delta,r} n^{-1/8})$.

This is easily seen to be a consequence of the following proposition, which makes explicit the tradeoff between the number r of rows removed and the small level sets parameter ε.

Proposition 5.1. *There is a constant $K_\delta > 0$ depending only on δ such that for any $\alpha \in (0, 1]$, $\varepsilon \in [K_\delta n^{-\alpha/8}, 0.9]$ and any $S \subset [n]$ with $|S| \leq (1 - \alpha)\varepsilon n$, with overwhelming probability any nontrivial right null vector of M_S has* SLS(ε).

Recall that a vector $x \in \mathbf{R}^n$ is said to be k-sparse if $|\operatorname{spt}(x)| \leq k$, where

$$\begin{aligned} \operatorname{spt}(x) &= \{i \in [n] : x(i) \neq 0\} \\ &= [n] \setminus x^{-1}(0). \end{aligned}$$

We start by showing that

(1) by row exchangeability, it suffices to consider the matrix M_r obtained by removing the last r rows, and
(2) it suffices to control the event that there are sparse vectors mapped by M_r to the subspace of constant vectors in $\mathbf{R}^{n-r}$.

This will reduce our task to proving the following:

Proposition 5.2. *For $r \in [n]$, denote by M_r the matrix $M_{[n-r]\times[n]}$ obtained from M by removing the last r rows. There is a constant $K_\delta > 0$ depending only on δ such that for any $\alpha \in (0, 1]$, $\varepsilon \in [K_\delta n^{-\alpha/8}, 0.9]$ and any $r \leq (1 - \alpha)\varepsilon n$, with overwhelming probability, there is no nonzero $(1 - \varepsilon)n$-sparse vector $x \in \mathbf{R}^n$ with $M_r x \in$* span($\mathbf{1}$).

Proof of Proposition 5.1. Let K_δ as in Proposition 5.2, and fix $\alpha \in (0, 1]$, $\varepsilon \in [K_\delta n^{-\alpha/8}, 0.9]$, and $r \in [\lfloor (1 - \alpha)\varepsilon n \rfloor]$. By row exchangeability it suffices to show that any nontrivial right null vector of $M_r := M_{[n-r]\times[n]}$ has SLS(ε).

Let $\mathcal{B}$ denote the event that there is a nonzero vector x with $|\operatorname{spt}(x)| \leq (1-\varepsilon)n$ such that $M_r x \in \operatorname{span}(\mathbf{1})$. By Proposition 5.2, $\mathcal{B}^c$ holds with overwhelming probability. Suppose that M_r has a nontrivial right null vector y such that such that $|y^{-1}(\lambda)| > \varepsilon n$ for some $\lambda \in \mathbf{R}$. Note that y is necessarily non-constant, as M_r cannot have constant nonzero null vectors. Let $z = \lambda \mathbf{1} - y$. Since y is non-constant, $z \neq 0$. Furthermore, $|\operatorname{spt}(z)| \leq (1 - \varepsilon)n$, and $Mz = \lambda d \mathbf{1} \in \operatorname{span}(\mathbf{1})$. Hence we are in event $\mathcal{B}$, and the claim follows. □

The remainder of this section is spent proving Proposition 5.2. The reader may wish to review the proof of Proposition 2.3 from Section 2 first, as it is a cartoon of the more complicated argument below.

We may assume by Proposition 3.1 and Theorem 3.5 that M has codegree and edge discrepancy properties. That is, for any $\lambda \in (0,1)$ independent of n, we have

$$h_{(i_1,i_2)}(M) \quad \text{and} \quad h_{(i_1,i_2)}(M^{\mathsf{T}}) \in (1-\lambda, 1+\lambda)v_\delta n \tag{5.1}$$

for all pairs of distinct indices $i_1, i_2 \in [n]$, and

$$e_{S,T}(M) \in (1-\lambda, 1+\lambda)\delta |S||T| \tag{5.2}$$

for all pairs of sets $S, T \subset [n]$ such that

$$|S||T| \geq K_{\lambda,\delta} n \tag{5.3}$$

for some $K_{\lambda,\delta} > 0$ depending only on λ and δ. We will later take λ sufficiently small depending on δ.

5.1. **Preliminary reductions.** Fix $\alpha \in (0,1]$, $\varepsilon \in [Kn^{-\gamma}, 0.9]$, and $r \in [\lfloor (1-\alpha)\varepsilon n \rfloor]$, for some $K, \gamma > 0$ to be determined.

Define the events

$$\mathcal{E}_k = \{\exists \text{ nontrivial } x \in \mathbf{R}^n : x \text{ is } k\text{-sparse, } M_r x \in \operatorname{span}(\mathbf{1})\}.$$

Our goal is to bound

$$\mathbf{P}\big(\mathcal{E}_{(1-\varepsilon)n}\big) = \sum_{k=2}^{\lfloor (1-\varepsilon)n \rfloor} \mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \tag{5.4}$$

(Note that $\mathcal{E}_1$ is empty, since no single column can be parallel to $\mathbf{1}$.)

For each $k \in [2, \lfloor (1-\varepsilon)n \rfloor]$, we bound $\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ by giving up some large combinatorial factors in order to pass to smaller events $\mathcal{E}_k'$ and $\mathcal{E}_k''$ (defined below) on which we can apply Theorem 2.2 to several independent random walks.

By column exchangeability and a union bound, we have

$$\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \leq \binom{n}{k} \mathbf{P}\big(\mathcal{E}_{[k]} \setminus \mathcal{E}_{k-1}\big) \tag{5.5}$$

where

$$\mathcal{E}_{[k]} = \{\exists x \in \mathbf{R}^n : \ \operatorname{spt}(x) = [k], \ M_r x \in \operatorname{span}(\mathbf{1})\}.$$

Note that by dilating x we have $\mathcal{E}_{[k]} = \mathcal{E}_{[k]}^0 \vee \mathcal{E}_{[k]}^1$, where

$$\mathcal{E}_{[k]}^0 = \{\exists x \in \mathbf{R}^n : \ \operatorname{spt}(x) = [k], \ M_r x = 0\}$$

and

$$\mathcal{E}_{[k]}^1 = \{\exists x \in \mathbf{R}^n : \ \operatorname{spt}(x) = [k], \ M_r x = \mathbf{1}\}.$$

From

$$\mathcal{E}_{[k]} = \mathcal{E}_{[k]}^0 \vee (\mathcal{E}_{[k]}^1 \setminus \mathcal{E}_{[k]}^0)$$

we may bound

$$\mathbf{P}\big(\mathcal{E}_{[k]} \setminus \mathcal{E}_{k-1}\big) \leq \mathbf{P}\Big(\mathcal{E}_{[k]}^0 \setminus \mathcal{E}_{k-1}\Big) + \mathbf{P}\Big(\mathcal{E}_{[k]}^1 \setminus (\mathcal{E}_{[k]}^0 \vee \mathcal{E}_{k-1})\Big). \tag{5.6}$$

For the first term on the right hand side, note that on $\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}$ the minor $M_{[n-r]\times[k]}$ has $k-1$ linearly independent rows. Indeed, if this were not the case we would have $\mathrm{rank}(M_{[n-r]\times[k]}) \le k-2$, so that $M_{[n-r]\times[k]}$ has 2 linearly independent right null vectors $\tilde{x}_1, \tilde{x}_2 \in \mathbf{R}^k$. But there is a $k-1$-sparse linear combination of $\tilde{x}_1, \tilde{x}_2$, putting us in $\mathcal{E}_{k-1}$.

For the second term in (5.6), note that on the complement of $\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1}$ the minor $M_{[n-r]\times[k]}$ has full rank, and hence has k linearly independent rows.

Now we spend some symmetry to fix the linearly independent rows. Let $\mathcal{L}_i$ denote the event that $R_1, \dots, R_i$ are linearly independent. By row exchangeability we have

$$\mathbf{P}\Big(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}\Big) \le \binom{n-r}{k-1} \mathbf{P}\Big(\Big(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}\Big) \wedge \mathcal{L}_{k-1}\Big) \tag{5.7}$$

and

$$\mathbf{P}\Big(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})\Big) \le \binom{n-r}{k} \mathbf{P}\Big(\Big(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})\Big) \wedge \mathcal{L}_k\Big). \tag{5.8}$$

In (5.7), $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ is the event that the first $k-1$ rows of M are linearly independent, that there is a null vector x of M_r supported on $[k]$, and that there are no $k-1$-sparse null vectors of M_r. Now on this event there is actually only one possibility for x up to dilation. Indeed, on $\mathcal{L}_{k-1}$ the system

$$M_{[k-1]\times[k]} y = 0 \tag{5.9}$$

has a unique solution up to dilation, by the linear independence of the first $k-1$ rows. Let us pick a nontrivial solution $\tilde{x} \in \mathbf{R}^k$ of (5.9) arbitrarily, and set $x^* = (\tilde{x} \quad 0)^\mathsf{T} \in \mathbf{R}^n$. On the complement of $\mathcal{E}_{k-1}$, each component of $\tilde{x}$ is nonzero. Hence, $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ is contained in the event

$$\begin{aligned} \mathcal{E}'_k &:= \mathcal{L}_{k-1} \wedge \{M_r x^* = 0\} \wedge \{\tilde{x}(i) \ne 0 \text{ for all } i \in [k]\} \\ &= \mathcal{L}_{k-1} \wedge \big\{M_{[k,n-r]\times[k]} \tilde{x} = 0\big\} \wedge \{\tilde{x}(i) \ne 0 \text{ for all } i \in [k]\}. \end{aligned}$$

We emphasize that $\tilde{x}$ is a random vector in $\mathbf{R}^k$, defined only on the event $\mathcal{L}_{k-1}$, determined by the minor $M_{[k-1]\times[k]}$ through (5.9). Also, the minor

$$M_{[k,n-r]\times[k]}$$

has at least $\lfloor \varepsilon n/2 \rfloor$ rows since $k \le (1-\varepsilon)n$ and $r \le \varepsilon n/2$.

We may similarly fix the vector in the preimage of $\mathbf{1}$ on the event $(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})) \wedge \mathcal{L}_k$ from (5.8). This event is disjoint from the event $(\mathcal{E}^0_{[k]} \setminus \mathcal{E}_{k-1}) \wedge \mathcal{L}_{k-1}$ from (5.7), and on it we may define $\tilde{y} \in \mathbf{R}^k$ as the unique solution of

$$M_{[k]\times[k]} y = \mathbf{1}. \tag{5.10}$$

Setting

$$\mathcal{E}''_k := \mathcal{L}_k \wedge \big\{M_{[k+1,n-r]\times[k]} \tilde{y} = \mathbf{1}\big\} \wedge \{\tilde{y}(i) \ne 0 \text{ for all } i \in [k]\} \tag{5.11}$$

we similarly conclude that

$$(\mathcal{E}^1_{[k]} \setminus (\mathcal{E}^0_{[k]} \vee \mathcal{E}_{k-1})) \wedge \mathcal{L}_k \subset \mathcal{E}''_k.$$

Here also, $\tilde{y} \in \mathbf{R}^k$ is a random vector defined only on the event $\mathcal{L}_k$ via (5.10).

Combined with (5.7), (5.8), (5.6) and (5.5), we have

$$\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \le \binom{n}{k}\binom{n-r}{k-1}\mathbf{P}(\mathcal{E}_k') + \binom{n}{k}\binom{n-r}{k}\mathbf{P}(\mathcal{E}_k''). \tag{5.12}$$

It remains to bound $\mathbf{P}(\mathcal{E}_k')$ and $\mathbf{P}(\mathcal{E}_k'')$. Letting $\varepsilon_0 > 0$ be a small constant (possibly depending on δ) to be chosen later, we will separately handle the cases of $k \in [2, \varepsilon_0 n]$, $k \in [\varepsilon_0 n, (1-\varepsilon_0)n]$ and $k \in [(1-\varepsilon_0)n, (1-\varepsilon)n]$.

5.2. **Bounding $\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ for small values of k.** Fix $k \in [2, \varepsilon_0 n]$. We construct a matrix $\widetilde{M}$ coupled and identically distributed to M using the switchings construction of Corollary 3.13 as follows. Let $m = \lfloor \lambda v_\delta n \rfloor$, and draw the following random variables, independently of each other and of M:

(1) $\sigma \in \mathrm{Sym}(n-k)$ a uniform random permutation on $n-k$ labels,
(2) $\xi = (\xi_l)_{l=1}^m$ a sequence of i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables, and
(3) $(I_l)_{l=1}^{2m}$, sampled uniformly without replacement from $[k+1, n-r]$. For each $l \in [m]$ we set $\mathbf{p}_l = (I_l^+, I_l^-) := (I_{2l-1}, I_{2l})$, and denote

$$\mathcal{P} = (\mathbf{p}_l)_{l=1}^m.$$

We emphasize that M, σ, ξ, and $\mathcal{P}$ are all independent of each other. First, we form M^σ from M by permuting the rows $k+1, \dots, n$ according to σ. Then, conditional on M and σ, for each $l \in [m]$ we perform a random switching of the minor $M^\sigma_{\mathbf{p}_l \times (1,n)}$ according to ξ_l, and let $\widetilde{M}$ be the resulting matrix. It follows from row exchangeability and Corollary 3.13 (with $\mathbf{q}_l = (1,n)$ for all $l \in [m]$) that $\widetilde{M}$ is also uniformly distributed in $\mathcal{M}_d$.

For brevity let A_l denote the lth sampled 2×2 minor $M^\sigma_{\mathbf{p}_l \times (1,n)}$. In order to show that A_l is switchable for many $l \in [m]$, we first use codegree discrepancy and the randomness of σ to argue that

$$h_+^\sigma(M) = \left|\left\{ i \in [k+1, n-r] : M^\sigma_{i \times (1,n)} = (1 \quad 0) \right\}\right|$$

and

$$h_-^\sigma(M) = \left|\left\{ i \in [k+1, n-r] : M^\sigma_{i \times (1,n)} = (0 \quad 1) \right\}\right|$$

are large. By (5.1) we have

$$\left|\left\{ i \in [1, n] : M_{i \times (1,n)} = (1 \quad 0) \right\}\right| = h_{(1,n)}(M^\mathsf{T}) \ge (1-\lambda) v_\delta n$$

so

$$\left|\left\{ i \in [k+1, n] : M_{i \times (1,n)} = (1 \quad 0) \right\}\right| \ge ((1-\lambda)v_\delta - \varepsilon_0)\, n.$$

It follows that

$$\begin{aligned}\mathbf{E}_\sigma\, h_+^\sigma(M) &\ge \frac{n-k-r}{n-k}((1-\lambda)v_\delta - \varepsilon_0)n \\ &\ge (1-\varepsilon/2)((1-\lambda)v_\delta - \varepsilon_0)n \\ &\gg (1-2\lambda)v_\delta n\end{aligned}$$

where we have taken $\varepsilon_0 \le \lambda v_\delta$ (and recall $\varepsilon \le 0.9$). We extend this to a lower bound on $h_+^\sigma(M)$ using concentration of measure for the symmetric group, conditioning on M

and applying Theorem 3.20 with $N = n-k$, $F(\sigma) = -h_+^\sigma(M)$, and t a sufficiently small multiple of $\mathbf{E}_\sigma\, h_+^\sigma(M)$ to conclude

$$h_+^\sigma(M) \gg (1-2\lambda)v_\delta n$$

except on an event of size $O(\exp(-c_{\lambda,\delta}n))$. We obtain the same lower bound for $h_-^\sigma(M)$ by an identical argument, and from a union bound we have

$$\min(h_+^\sigma(M), h_-^\sigma(M)) \gg (1-2\lambda)v_\delta n \tag{5.13}$$

except on an event of size $O(\exp(-c_{\lambda,\delta}n))$.

Now condition on M, σ such that (5.13) holds. For each $l \in [m]$ we have

$$\begin{aligned}\mathbf{P}(A_l \text{ is switchable}) &\ge \mathbf{P}(A_l = \mathbf{I}_2)\\ &\ge \left(\frac{h_+^\sigma(M) - 2m}{n}\right)\left(\frac{h_-^\sigma(M) - 2m}{n}\right)\\ &\gg (1-4\lambda)^2 v_\delta^2\\ &\gg v_\delta^2\end{aligned}$$

taking λ sufficiently small, where the presence of $2m$ in the second line (over) compensates for the fact that $\{I_l\}_{l=1}^{2m}$ are being sampled without replacement. Now letting

$$L = \{l : A_l \text{ is switchable}\},$$

it follows from Lemma 3.22 that $|L| \gg_\delta n$ except on an event of size $O(\exp(-c_\delta n))$.

Since M and $\widetilde{M}$ are identically distributed, we have $\mathbf{P}(\mathcal{E}_k') = \mathbf{P}(\widetilde{\mathcal{E}}_k')$ and $\mathbf{P}(\mathcal{E}_k'') = \mathbf{P}(\widetilde{\mathcal{E}}_k'')$, where $\widetilde{\mathcal{E}}_k'$ and $\widetilde{\mathcal{E}}_k''$ are the events that $\mathcal{E}_k'$ and $\mathcal{E}_k''$ hold for $\widetilde{M}$ rather than M – that is,

$$\widetilde{\mathcal{E}}_k' := \mathcal{L}_{k-1} \wedge \left\{\widetilde{M}_{[k,n-r]\times[k]}\tilde{x} = 0\right\} \wedge \{\tilde{x}(i) \neq 0 \text{ for all } i \in [k]\}$$

and

$$\widetilde{\mathcal{E}}_k'' := \mathcal{L}_k \wedge \left\{\widetilde{M}_{[k+1,n-r]\times[k]}\tilde{y} = \mathbf{1}\right\} \wedge \{\tilde{y}(i) \neq 0 \text{ for all } i \in [k]\}.$$

(Note that since M and $\widetilde{M}$ are identical on the first k rows, they determine the same vectors $\tilde{x}$ and $\tilde{y}$.)

Let us first consider $\widetilde{\mathcal{E}}_k'$. For each $l \in [m]$ we have

$$\mathbf{P}\left(\widetilde{M}_{\mathbf{p}_l\times[k]}\tilde{x} = \begin{pmatrix}0\\0\end{pmatrix} \,\middle|\, M, \sigma, \mathcal{P}\right) = \mathbf{P}\left(\widetilde{M}_{\mathbf{p}_l\times 1}\tilde{x}(1) + M_{\mathbf{p}_l\times[2,k]}(\tilde{x})_{[2,k]} = \begin{pmatrix}0\\0\end{pmatrix} \,\middle|\, M, \sigma, \mathcal{P}\right).$$

Note that $\tilde{x}$ is deterministic under the conditioning on M. Also, for each l such that A_l is switchable, the term $M_{\mathbf{p}_l\times[2,k]}(\tilde{x})_{[2,k]}$ is deterministic under the conditioning on M and $\mathcal{P}$. For the first term we have

$$\widetilde{M}_{\mathbf{p}_l\times 1} = \begin{pmatrix}1\\0\end{pmatrix} 1_{\xi_l=+1} + \begin{pmatrix}0\\1\end{pmatrix} 1_{\xi_l=-1}$$

and since $\tilde{x}(1) \neq 0$ we have

$$\mathbf{P}\left(\widetilde{M}_{\mathbf{p}_l\times[k]}\tilde{x} = \begin{pmatrix}0\\0\end{pmatrix} \,\middle|\, M, \sigma, \mathcal{P}\right) \le \frac{1}{2} \tag{5.14}$$

for all l such that A_l is switchable. Hence we can bound

$$\begin{aligned}
\mathbf{P}(\mathcal{E}'_k) &= \mathbf{P}\Big(\widetilde{\mathcal{E}}'_k\Big) \\
&= \mathbf{E}\,\mathbf{P}\Big(\widetilde{\mathcal{E}}'_k\Big|M\Big) \\
&\le O(\exp(-c_\delta n)) + \mathbf{E}\,\mathbf{P}\Big(\widetilde{\mathcal{E}}'_k\Big|M,\sigma,\mathcal{P}\Big)1_{\{|L|\ge c_\delta n\}} \\
&\le O(\exp(-c_\delta n)) + \mathbf{E}\,\mathbf{P}\left(\widetilde{M}_{\mathbf{p}_l\times[k]}\tilde{x} = \begin{pmatrix}0\\0\end{pmatrix} \text{ for all } l\in[m] \,\middle|\, M,\sigma,\mathcal{P}\right)1_{\{|L|\ge c_\delta n\}} \qquad (5.15)\\
&\le O(\exp(-c_\delta n)) + 2^{-|L|}1_{\{|L|\ge c_\delta n\}} \\
&= O(\exp(-c_\delta n))
\end{aligned}$$

for some altered value of c_δ, where the second to last line follows from (5.14) and independence of the $(\xi_l)_{l=1}^m$.

We similarly obtain

$$\mathbf{P}(\mathcal{E}''_k) = O(\exp(-c_\delta n))$$

by replacing $\binom{0}{0}$ with $\binom{1}{1}$ in (5.15), and $\tilde{x}$ with $\tilde{y}$. Together with (5.12) we conclude that for $k \in [2, \varepsilon_0 n]$,

$$\begin{aligned}
\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) &\le \binom{n}{k}^2 O(\exp(-c_\delta n)) \\
&\le \left(\frac{e}{\varepsilon_0}\right)^{2\varepsilon_0 n} O(\exp(-c_\delta n)) \\
&\ll \exp(n(2\varepsilon_0(1-\log\varepsilon_0) - c_\delta)) \\
&\ll \exp(-c'_\delta n) \qquad (5.16)
\end{aligned}$$

for ε_0 sufficiently small depending on δ.

5.3. Bounding $\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ for medium values of k.

Here we bound $\mathbf{P}(\mathcal{E}'_k)$ and $\mathbf{P}(\mathcal{E}''_k)$ for $k \in [\varepsilon_0 n, (1-\max(\varepsilon_0,\varepsilon))n]$. If $\varepsilon \ge \varepsilon_0$ then this is the only remaining case, but recall we are allowing ε to be as small as $Kn^{-\gamma}$. Our aim is to replace M with a certain coupled matrix $\widetilde{M}$, in terms of which we can express the singularity events

$$\Big\{\widetilde{M}_{[k,n-r]\times[k]}\tilde{x} = 0\Big\}, \quad \Big\{\widetilde{M}_{[k,n-r]\times[k]}\tilde{y} = \mathbf{1}\Big\}$$

from $\mathcal{E}'_k, \mathcal{E}''_k$ as the events that several independent random walks end at 0 and 1, respectively.

Recall from Komlós' proof for the i.i.d. sign matrix B (see Proposition 2.3) we had

$$\begin{aligned}
\mathbf{P}\big(B_{[k,n]\times[k]}(x)_{[k]} = 0\big) &= \mathbf{P}(R_i\cdot x = 0 \quad \forall k\le i\le n) \\
&= \mathbf{P}(R_n\cdot x = 0)^{n-k+1}
\end{aligned}$$

where we have used independence of the rows R_i in the second line. The random variables $R_n \cdot x$ are random walks with k nonzero steps, to which we applied Theorem 2.2.

To replicate this for M, we will form $\widetilde{M}$ by applying several independent switchings on several pairs of rows. Conditional on M, this will render the product

$$\widetilde{M}_{[k,n-r]\times[k]}\tilde{x}$$

as a vector of independent random walks:

$$\widetilde{M}_{(i_1,i_2)\times[k]}\tilde{x} \in \mathbf{R}^2,$$

one for each pair of rows. Now for each pair (i_1, i_2) we will want to apply Theorem 2.2 to the i_1 component, say, to control the event that this (scalar) random walk lands at 0. As in the high level proof of Theorem 1.2 from Section 4, the steps of this walk will involve differences of components of x evaluated at sampled column indices:

$$x(J_l^+) - x(J_l^-).$$

In the present setting, we can ensure that many of these steps are nonzero by sampling pairs (J_l^+, J_l^-) uniformly with $J_l^+ \in [k]$ and $J_l^- \in [k+1, n]$, since we have already reduced to the case that $x(j) = 0$ for $j \in [k+1, n]$ and $x(j) = \tilde{x}(j) \neq 0$ for $j \in [k]$.

However, we will need to show that with this method of sampling the minor $M_{(i_1,i_2)\times(J_l^+,J_l^-)}$ is likely to be switchable. In Section 4 we could do this using the codegree discrepancy property (Proposition 3.1), since we were sampling J_l^+, J_l^- from the full range $[n]$. Here we need

$$M_{(i_1,i_2)\times j_1} = \begin{pmatrix}1\\0\end{pmatrix}, \quad M_{(i_1,i_2)\times j_2} = \begin{pmatrix}0\\1\end{pmatrix}$$

for many indices j_1, j_2 in *both* $[k]$ and $[k+1, n]$.

We phrase this more precisely as follows. For an ordered pair of rows (i_1, i_2) and a subset of column indices $T \subset [n]$, we define the statistic

$$h_{(i_1,i_2)\times T}(M) := \left| \left\{ j \in T : M_{(i_1,i_2)\times j} = \begin{pmatrix}1\\0\end{pmatrix} \right\} \right|. \tag{5.17}$$

This relates to our previous notation via

$$h_{(i_1,i_2)}(M) = h_{(i_1,i_2)\times[n]}(M).$$

While Proposition 3.1 tells us that with overwhelming probability

$$h_{(i_1,i_2)\times[n]}(M) \approx v_\delta n$$

for every pair (i_1, i_2), we want

$$h_{(i_1,i_2)\times T}(M) \approx v_\delta |T| \tag{5.18}$$

for every pair (i_1, i_2) and (at least) for all T of the form $[k]$ or $[k+1, n]$ with $k \in [\varepsilon_0 n, (1-\varepsilon)n]$. See Figure 2.

The following lemma tells us that the ostensibly stronger codegree discrepancy property (5.18) essentially already follows from the codegree and edge discrepancy properties, where the caveat is that we must pass to a large subset of pairs (i_1, i_2).

Lemma 5.3. *Let $A \in \mathcal{M}_d$ and suppose that for some $\lambda > 0$ sufficiently small depending on δ and $K > 0$, A has codegree and edge discrepancy properties: for all distinct $i_1, i_2 \in [n]$,*

$$h_{(i_1,i_2)\times[n]}(A) \quad \text{and} \quad h_{(i_1,i_2)\times[n]}(A^{\mathsf{T}}) \in (1-\lambda, 1+\lambda)v_\delta n \tag{5.19}$$

and for all pairs $S_0, T_0 \subset [n]$ such that $|S_0||T_0| \geq Kn$ we have

$$e_{S_0,T_0}(A) \in (1-\lambda, 1+\lambda)\delta|S_0||T_0|. \tag{5.20}$$

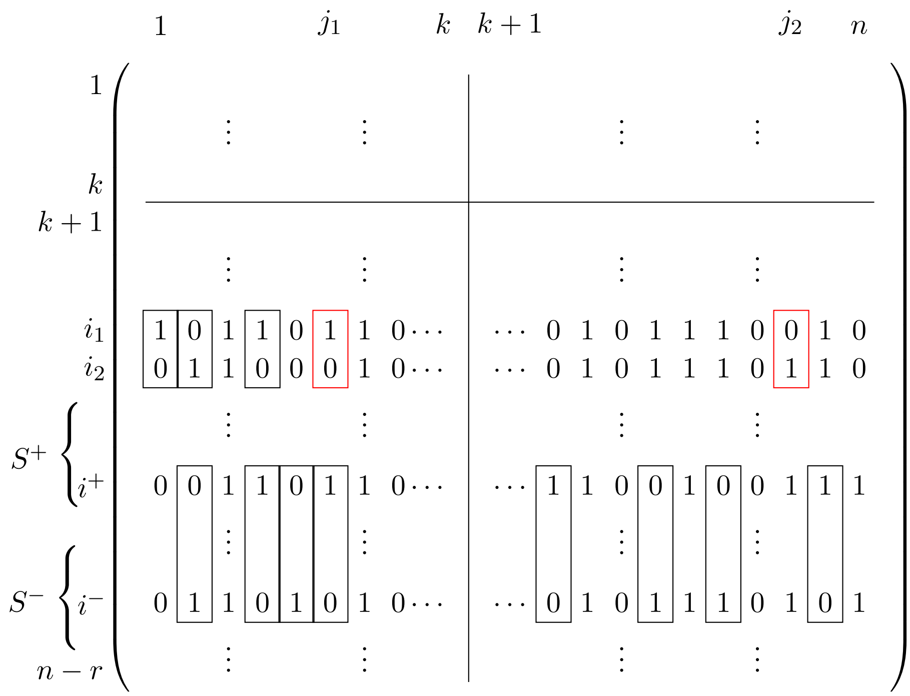

FIGURE 2. As in the proof of Proposition 2.3, we seek to apply Erdős' anti-concentration estimate to several independent random walks. For this to be effective, we need to find pairs of row indices (i_1, i_2) on which we can sample switchable 2×2 minors $M_{(i_1,i_2)\times(j_1,j_2)}$ with $j_1 \in [k]$ and $j_2 \in [k+1, n]$, as with the minor boxed in red above. This in turn requires that there be many 2×1 minors equal to $\binom{1}{0}$ and $\binom{0}{1}$ on *both* sides of the partition $[k] \cup [k+1, n]$. While the pair (i_1, i_2) depicted above has many such minors with columns in the range $[k]$, there are not many partners for them in the range $[k+1, n]$. This problem is resolved with Lemma 5.3, which gives us large sets S^+, S^- of row indices such that most pairs $(i^+, i^-) \in S^+ \times S^-$ have roughly the expected number of minors equal to $\binom{1}{0}, \binom{0}{1}$ on both sides of the partition.

Then for any fixed $S, T \subset [n]$ with $|S||T| \geq \frac{10K}{\lambda} n$, there exist disjoint subsets $S^+, S^- \subset S$ with

$$|S^+| = |S^-| \gg \lambda^2 \delta |S|$$

such that for each $i^+ \in S^+$, we have

$$h_{(i^+,i^-)\times T}(A) = \big(1 + O(\lambda)\big) v_\delta |T| \tag{5.21}$$

$$h_{(i^-,i^+)\times T}(A) = \big(1 + O(\lambda)\big) v_\delta |T| \tag{5.22}$$

$$h_{(i^+,i^-)\times [n]\setminus T}(A) = \big(1 + O(\lambda)\big) v_\delta (n - |T|) \tag{5.23}$$

$$h_{(i^-,i^+)\times [n]\setminus T}(A) = \big(1 + O(\lambda)\big) v_\delta (n - |T|) \tag{5.24}$$

for all but $O\big(\frac{Kn}{v_\delta |T|}\big)$ elements i^- of S^-.

Remark 5.4. The lemma is in some sense optimal, in that passing to subsets S^+, S^- should be necessary. Indeed, one could prove a slightly generalized codegree discrepancy

property for the random variables $h_{(i_1,i_2)\times[k]}(M)$ for a range of $k \in [n]$, but the failure bound for this property would be of size $\exp(-c_{\lambda,\delta}n)$, which is optimal, and does not beat the binomial coefficients from the union bounds we took in order to reduce to the case that $T = [k]$. Hence we have taken the approach of establishing the codegree and edge discrepancy properties – which do not depend on row and column labelings – outside the main argument, and applying deterministic consequences of these inside the main argument.

A discrepancy property for the random variables $h_{(i_1,i_2)\times T}(M)$ for arbitrary sufficiently large $T \subset [n]$ would be independent of row and column labelings, but is obviously false (just take T to be the set of indices j where $M_{(i_1,i_2)\times j} = \binom{1}{0}$).

We defer the proof of this lemma to the appendix. The idea is essentially that with the codegree discrepancy property (5.19), in order for a pair of rows (i^+, i^-) to fail estimates (5.21 – 5.24) there will need to be a lot of minors of the form $\binom{1}{1}, \binom{0}{0}$ on one side of the partition, and if this happens for many values of i^- we can locate a minor $A_{S'\times T'}$ that is either more dense or more sparse than the edge discrepancy property (5.20) allows.

The conclusion of the lemma holds for M if we take $K = K_{\lambda,\delta}$ from (5.20). Call a pair of indices (i^+, i^-) "λ-good" if we have (5.21 – 5.24) with $A = M$ and $T = [k+1, n]$. We apply Lemma 5.3 with $S = [k+1, n-r], T = [k+1, n]$ to get disjoint subsets S^+, S^- of $[k+1, n-r]$ with

$$s_\pm := |S^+| = |S^-| \gg_{\lambda,\delta} (n-k-r)$$

such that for each $i^+ \in S^+$ we have (i^+, i^-) is λ-good for all but $O_{\lambda,\delta}(n/(n-k))$ elements i^- of S^-. Since $k \in [\varepsilon_0 n, (1-\max(\varepsilon, \varepsilon_0))n]$ and $r \leq (1-\alpha)\varepsilon n$, it follows that

$$n - k - r \geq \alpha\varepsilon_0 n \gg_\delta n \tag{5.25}$$

where we have used that $\alpha \in (0, 1]$ is fixed and ε_0 depends only on δ. Hence,

$$s_\pm \gg_{\lambda,\delta} n. \tag{5.26}$$

With a union bound we may pay a factor 2^{2n} to fix S^+, S^- as deterministic sets. For concreteness say $S^+ = [k+1, k+s_\pm]$ and $S^- = [k+s_\pm+1, k+2s_\pm]$.

For each $i^+ \in S^+$, let $S(i^+)$ be the subset of S^- of indices i^- such that one of (5.21 – 5.24) fails. We have

$$|S(i^+)| \leq C_{\lambda,\delta}\frac{n}{n-k}$$

for all $i^+ \in S^+$.

Let $\sigma : S^+ \to S^-$ be a uniform random bijection. Conditional on M, define the bad event

$$\mathcal{B}_\sigma = \{\text{for} \geq 0.1 s_\pm \text{ values of } i \in S^+, \sigma(i) \in S(i)\}.$$

By the bound on $|S(i^+)|$ we have

$$\begin{aligned}
\mathbf{P}_\sigma(\mathcal{B}_\sigma) &\le \binom{s_\pm}{0.1s_\pm}\frac{(C_{\lambda,\delta}n/(n-k))^{0.1s_\pm}(0.9s_\pm)!}{s_\pm!} \\
&\le 2^{s_\pm}\frac{C_{\lambda,\delta}^{0.1s_\pm}(n/(n-k))^{0.1s_\pm}(0.9s_\pm)^{0.9s_\pm}}{(s_\pm/e)^{s_\pm}} \\
&\le \exp\left(s_\pm\left(C'_{\lambda,\delta}-0.1\left(\log\left(\frac{n-k}{n}\right)+\log s_\pm\right)\right)\right) \\
&\le \exp\left(c'_{\lambda,\delta}n\left(C''_{\lambda,\delta}-0.1\log n\right)\right) \\
&\ll_{\lambda,\delta} \exp\left(-c_{\lambda,\delta}n\log n\right) \qquad (5.27)
\end{aligned}$$

where we have used (5.26) in the penultimate line. By row-exchangeability it follows that for any fixed $\sigma : S^+ \to S^-$, say $\sigma := i \mapsto i + s_\pm$ for concreteness, we have $\sigma(i) \notin S(i)$ for at least $0.9s_\pm$ values of $i \in S^+$, except on an event (now in the randomness of M) of size bounded by the expression in (5.27). We can assume that these good elements of S^+ are

$$[k+1, k+0.9s_\pm] =: S^{++}$$

by paying a factor of $\binom{s_\pm}{0.9s_\pm}$, which we can crudely bound by 2^n.

We have now fixed a set of λ-good pairs of rows with indices $\{(i,\sigma(i))\}_{i\in S^{++}}$. Next we randomly sample pairs of column indices. We abbreviate $k^* := \min(k, n-k)$, and set $m = \lfloor \varepsilon_1 k^*\rfloor$ with $\varepsilon_1 > 0$ to be chosen later. For each $i \in S^{++}$, let

$$\mathcal{Q}_i = \{\mathbf{q}_{i,l}\}_{l=1}^m := \{(J_{i,l}^+, J_{i,l}^-)\}_{l=1}^m \qquad (5.28)$$

where the samples $\{\mathcal{Q}_i\}_{i\in S^{++}}$ are jointly independent, and all independent of M, with $\mathbf{q}_1^i$ sampled uniformly from $[k] \times [k+1,n]$, and for each $l \in [2,m]$ $\mathbf{q}_{i,l}$ is sampled uniformly from

$$\left([k]\setminus\{J_{i,l'}^+\}_{l'=1}^{l-1}\right)\times\left([k+1,n]\setminus\{J_{i,l'}^-\}_{l'=1}^{l-1}\right).$$

For brevity we denote by $A_{i,l}$ the randomly sampled 2×2 minor $M_{(i,\sigma(i))\times\mathbf{q}_{i,l}}$. For any $i \in S^{++}$, since $(i,\sigma(i))$ is λ-good we have

$$\mathbf{P}(A_{i,1} = \mathbf{I}_2) = (1+O_\delta(\lambda))v_\delta^2 \qquad (5.29)$$

and

$$\mathbf{P}(A_{i,1} = \mathbf{J}_2) = (1+O_\delta(\lambda))v_\delta^2. \qquad (5.30)$$

Furthermore, for each $l \in [m]$ we can lower bound

$$\begin{aligned}
\mathbf{P}(A_{i,l} \text{ is switchable}) &\ge (2+O_\delta(\lambda))v_\delta^2 - \frac{l-1}{k^*} \\
&\ge (2+O_\delta(\lambda))v_\delta^2 - \varepsilon_1 \\
&\gg_\delta 1
\end{aligned}$$

for λ and ε_1 sufficiently small depending on δ.

Letting

$$L_i := \{l \in [m] : A_{i,l} \text{ is switchable}\}$$

we have by Lemma 3.22 that for fixed $i \in S^{++}$

$$|L_i| \gg_\delta k^* \gg_\delta n$$

except on an event of size at most

$$C\exp(-c_\delta n) =: p.$$

This is not small enough to conclude that $|L_i| \gg_\delta n$ for all $i \in S^{++}$ using a union bound, since the failure probability has to beat the binomial coefficients from (5.12). However, it will be enough to have $|L_i| \gg_\delta n$ for *most* $i \in S^{++}$.

Let

$$U = \sum_{i\in S^{++}} 1_{\{|L_i|<c_\delta n\}}$$

for c_δ sufficiently small, and define the bad event

$$\mathcal{B}_{\mathcal{Q}} := \{U > s_\pm/2\}. \tag{5.31}$$

U is binomially distributed by the independence of the $\mathcal{Q}_i$. We can crudely bound the tail of U as follows:

$$\begin{aligned}\mathbf{P}(\mathcal{B}_{\mathcal{Q}}) &\le \sum_{s=s_\pm/2}^{0.9s_\pm} \binom{0.9s_\pm}{s} p^s(1-p)^{0.9s_\pm - s}\\ &\le s_\pm 2^{s_\pm} p^{s_\pm/2}\\ &= O(\exp(cs_\pm \log p))\\ &\ll_\delta \exp\left(-c_\delta n^2\right). \end{aligned}\tag{5.32}$$

On the complement of this event, $|L_i| \ge c_\delta n$ for at least $s_\pm/2$ elements of S^{++}. We spend another factor of 2^n to assume $|L_i| \ge c_\delta n$ for all $i \in [k+1, k+s_\pm/2] =: S^*$.

Let us summarize our progress so far. We have

$$\begin{aligned}\mathbf{P}(\mathcal{E}_k') &\le 2^{2n}\left(\mathbf{P}_\sigma(\mathcal{B}_\sigma) + 2^n\big(\mathbf{P}(\mathcal{B}_{\mathcal{Q}}) + 2^n\,\mathbf{P}(\mathcal{E}_k' \wedge \mathcal{G})\big)\right)\\ &\le 2^{4n}\left(\mathbf{P}_\sigma(\mathcal{B}_\sigma) + \mathbf{P}(\mathcal{B}_{\mathcal{Q}}) + \mathbf{P}(\mathcal{E}_k' \wedge \mathcal{G})\right)\end{aligned}\tag{5.33}$$

and similarly for $\mathcal{E}_k''$, where

$$\mathcal{G} := \{\forall i \in S^*,\ |L_i| \ge c_\delta n\}. \tag{5.34}$$

We are now in a position to define a powerful enough coupling $(M, \widetilde{M})$. Let $\{\xi_{i,l}\}_{i\in S^*, l\in[m]}$ be i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables, independent of all other random variables in play, and conditional on M form $\widetilde{M}$ by applying random switchings according to the $\xi_{i,l}$ to the sampled minors $A_{i,l}$ for all $i \in S^*, l \in [m]$. By Corollary 3.13, $\widetilde{M}$ is also uniformly distributed on $\mathcal{M}_d$.

We have

$$\begin{aligned}\mathbf{P}\big(\mathcal{E}_k' \wedge \mathcal{G}\big) &= \mathbf{P}\Big(\widetilde{\mathcal{E}}_k' \wedge \mathcal{G}\Big)\\ &= \mathbf{E}\,\mathbf{P}\Big(\widetilde{\mathcal{E}}_k' \Big| M, \{\mathcal{Q}_i\}_{i\in S^*}\Big)1_{\mathcal{G}}\\ &\le \mathbf{E}\,\mathbf{P}\Big(\Big\{\widetilde{M}_{S^*\times[k]}\tilde{x} = 0\Big\}\Big| M, \{\mathcal{Q}_i\}_{i\in S^*}\Big)1_{\mathcal{G}}\\ &= \mathbf{E}\prod_{i\in S^*}\mathbf{P}\Big(\Big\{\widetilde{M}_{i\times[k]}\tilde{x} = 0\Big\}\Big| M, \mathcal{Q}_i\Big)1_{\mathcal{G}}\end{aligned}\tag{5.35}$$

where we have used the independence of the $\xi_{i,l}$ in the last line. We may express $\widetilde{M}_{i\times[k]}\tilde{x}$ as a random walk:

$$\begin{aligned}\widetilde{M}_{i\times[k]}\tilde{x} &= s_{i,0}(x) + \sum_{l\in L_i} \widetilde{M}(i, J^+_{i,l})\tilde{x}(J^+_{i,l}) \\ &= s_{i,0}(x) + \frac{1}{2}\sum_{l\in L_i} \tilde{x}(J^+_{i,l}) + \frac{1}{2}\sum_{l\in L_i} \xi_{i,l}\tilde{x}(J^+_{i,l})\end{aligned}$$

where $s_{i,0}(x)$ collects contributions to the sum $\widetilde{M}_{i\times[k]}\tilde{x}$ from unsampled indices. Since $\tilde{x}(j)\neq 0$ for all $j\in[k]$ on the event $\mathcal{E}'_k$, we conclude from Theorem 2.2 that

$$\begin{aligned}\mathbf{P}\Big(\Big\{\widetilde{M}_{i\times[k]}\tilde{x}=0\Big\}\Big| M, \mathcal{Q}_i\Big)1_{\mathcal{G}} &\ll |L_i|^{-1/2}1_{\mathcal{G}} \\ &\ll (c_\delta n)^{-1/2}\end{aligned}$$

where the second inequality follows from our restriction to the event $\mathcal{G}$. Inserting this in (5.35),

$$\mathbf{P}(\mathcal{E}'_k\wedge\mathcal{G}) \le O\bigg(\frac{1}{\sqrt{c_\delta n}}\bigg)^{|S^*|} = O_\delta\bigg(\frac{1}{\sqrt{n}}\bigg)^{c'_\delta n}. \tag{5.36}$$

We obtain the same bound for $\mathbf{P}(\mathcal{E}''_k\wedge\mathcal{G})$, mutatis mutandis.

Now we combine all of our bounds. From (5.33), (5.27), (5.32) and (5.36) we have

$$\begin{aligned}\mathbf{P}\big(\mathcal{E}'_k\big) &\ll_\delta 2^{4n}\big(\exp(-c_\delta n\log n) + \exp(-c_\delta n^2) + \exp(-c_\delta n\log n)\big) \\ &\ll_\delta \exp(-c'_\delta n\log n)\end{aligned}$$

and we obtain the same bound for $\mathbf{P}(\mathcal{E}''_k)$. Together with (5.12) (again bounding binomial coefficients by 2^n) we conclude

$$\begin{aligned}\mathbf{P}(\mathcal{E}_k\setminus\mathcal{E}_{k-1}) &\le 2^{2n}(\mathbf{P}\big(\mathcal{E}'_k\big) + \mathbf{P}\big(\mathcal{E}''_k\big)) \\ &\ll_\delta \exp(-c_\delta n\log n).\end{aligned} \tag{5.37}$$

5.4. **Bounding $\mathbf{P}(\mathcal{E}_k\setminus\mathcal{E}_{k-1})$ for large values of k.** In (5.16) we took ε_0 sufficiently small depending on δ. If $\varepsilon<\varepsilon_0$, it only remains to bound $\mathbf{P}(\mathcal{E}'_k)$ and $\mathbf{P}(\mathcal{E}''_k)$ for k in the range $[\varepsilon_0 n, (1-\varepsilon)n]$.

First we make a progress report: by summing our bounds (5.16) and (5.37) on $\mathbf{P}(\mathcal{E}_k\setminus\mathcal{E}_{k-1})$ for $k\in[2,(1-\varepsilon_0)n]$, we see that the conclusion of Proposition 5.1 already holds for $\varepsilon\in[\varepsilon_0, .9]$. In particular, we may now assume that any right null vector of M_r has $\mathrm{SLS}(\varepsilon_0)$. This implies that $\tilde{x}$ and $\tilde{y}$, fixed by the conditioning on $R_1,\dots,R_k$, have $\mathrm{SLS}(\varepsilon_0)$.

As we did for medium values of k, we will define a matrix $\widetilde{M}$ coupled to M using independent samples $\mathcal{Q}'_i$ of pairs of column indices. The difference here is that the observation that $\tilde{x}$ and $\tilde{y}$ have $\mathrm{SLS}(\varepsilon_0)$ will allow us to sample column indices from the full range of $[n]$ as in Section 4, rather than from either side of the partition $[k]\cup[k+1,n]$. In particular, we will not need to pass to a "good" subset of row-pairs as we did with Lemma 5.3.

We turn to the details. Since we have conditioned on the first k rows of M_r and $r\le(1-\alpha)\varepsilon n$, there are at least

$$n-k-r\ge\alpha(n-k)$$

free rows remaining. Let

$$S=[n-k+1, n-k+\lfloor\alpha(n-k)/2\rfloor]$$

and for each $i \in S$, let

$$\mathbf{p}_i = (i, i + \lfloor \alpha(n-k)/2 \rfloor).$$

Set $m = \lfloor \varepsilon_2 v_\delta n \rfloor$ with $\varepsilon_2 > 0$ to be chosen later. For each $i \in S$, sample $J_{i,1}, \dots, J_{i,2m}$ uniformly without replacement from $[n]$, where the $|S|$ samples are performed independently of each other and of M. Set $\mathbf{q}'_{i,l} = (J_{i,2l-1}, J_{i,2l})$ for each $l \in [m]$, and set

$$\mathcal{Q}'_i = (\mathbf{q}_{i,l})_{l=1}^m.$$

Let $\{\xi_{i,l}\}_{i \in S, l \in [m]}$ be i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables independent of all other variables in play. Conditional on M and $\mathcal{Q}'_i$ for each $i \in S$, form $\widetilde{M}$ by applying a random switching according to $\xi_{i,l}$ at $\mathbf{p}_i \times \mathbf{q}'_{i,l}$ for each $i \in S, l \in [m]$. $\widetilde{M}$ is identically distributed to M by Corollary 3.13.

Conditional on M and $\mathcal{Q}'_i$ we may express the random variable $\widetilde{M}_{i \times [n]} x$ a random walk:

$$\widetilde{M}_{i \times [n]} x = s'_{i,0}(x) + \sum_{l \in L'_i} a_{i,l}(x) + \xi_{i,l} \partial_{i,l}(x) \tag{5.38}$$

where

$$\begin{aligned} a_{i,l}(x) &= \frac{x(J_{2l-1}) + x(J_{2l})}{2} \\ \partial_{i,l} &= \frac{x(J_{2l-1}) - x(J_{2l})}{2} \end{aligned}$$

and $s'_{i,0}(x)$ collects the summands of $\widetilde{M}_{i \times [n]} x$ coming from unsampled indices.

Denote $A_{i,l} = M_{\mathbf{p}_i \times \mathbf{q}'_{i,l}}$. For any $i \in S$ and $l \in [m]$, it follows from the codegree discrepancy property (5.1) that

$$\begin{aligned} \mathbf{P}(A_{i,l} \text{ is switchable}) &\geq \left(\frac{h_{\mathbf{p}_i}(M) - 2m}{n} \right)^2 \\ &\geq ((1-\lambda) v_\delta - 2\varepsilon_2)^2 \\ &\gg v_\delta^2 \end{aligned}$$

for λ, ε_2 sufficiently small. Now letting

$$L'_i = \{ l \in [m] : A_{i,l} \text{ is switchable} \}$$

we have by Corollary 3.22 that for each $i \in S$, $|L'_i| \gg_\delta n$ except on an event of size $O(\exp(-c_\delta n)$. By a union bound we may restrict the good event $\mathcal{G}'$ that $|L'_i| \gg_\delta n$ for all $i \in S$.

Let

$$L^*_i = \left\{ j \in L'_i : x(J_{2l-1}) \neq x(J_{2l}) \right\}.$$

We may argue exactly as we did for the estimate (4.14) to conclude that for fixed $i \in S$,

$$|L^*_i| \gg_\delta n \tag{5.39}$$

except on an event of size $O(\exp(-c_\delta n))$ by taking ε_0 smaller, if necessary, depending on δ. By a union bound, we conclude that the event

$$\mathcal{G}^* := \left\{ |L^*_i| \geq c'_\delta n \ \ \forall i \in S \right\}$$

holds with probability $1 - O_\delta(\exp(-c_\delta n))$ for some $c'_\delta > 0$ sufficiently small.

As in (5.35) we obtain

$$\mathbf{P}\big(\mathcal{E}_k' \wedge \mathcal{G}^*\big) \leq \mathbf{E}\prod_{i\in S}\mathbf{P}\Big(\Big\{\widetilde{M}_{i\times[n]}x = 0\Big\}\,\big|M, \mathcal{Q}_i'\Big)1_{\mathcal{G}^*}. \tag{5.40}$$

Applying Theorem 2.2 to each term, and using the lower bound on $|L_i^*|$ we have

$$\mathbf{P}\big(\mathcal{E}_k' \wedge \mathcal{G}^*\big) = O\left(\frac{1}{\sqrt{c_\delta' n}}\right)^{|S|} = O_\delta\left(\frac{1}{\sqrt{n}}\right)^{\frac{\alpha}{2}(n-k)}. \tag{5.41}$$

By identical reasoning we obtain the same bound on $\mathbf{P}(\mathcal{E}_k'' \wedge \mathcal{G}^*)$.

We combine this estimates into (5.12) to get

$$\begin{aligned}\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) &\leq \binom{n}{n-k}^2 \left(\mathbf{P}(\mathcal{G}^{*c}) + \mathbf{P}\big(\mathcal{E}_k' \wedge \mathcal{G}^*\big) + \mathbf{P}\big(\mathcal{E}_k'' \wedge \mathcal{G}^*\big)\right)\\ &\leq \binom{n}{n-k}^2 \left(C_\delta \exp(-c_\delta n) + 2\left(\frac{C_\delta}{\sqrt{n}}\right)^{\frac{\alpha}{2}(n-k)}\right)\\ &= \mathrm{I} + \mathrm{II}.\end{aligned} \tag{5.42}$$

For the binomial coefficients we have

$$\binom{n}{n-k}^2 \leq \left(\frac{en}{n-k}\right)^{2(n-k)} \tag{5.43}$$

$$\begin{aligned}&= \exp\left(2(n-k)\big(1+\log\frac{n}{n-k}\big)\right)\\ &\leq \exp\left(2\varepsilon_0\left(1-\log\varepsilon_0\right)n\right)\end{aligned} \tag{5.44}$$

where we have used the bound $n-k \leq \varepsilon_0 n$ and the fact that $G(x) = -x\log x$ is increasing on $[0, 1/e)$. Using (5.44) we conclude

$$\mathrm{I} \ll_\delta \exp\left(-c_\delta n\right) \tag{5.45}$$

by taking ε_0 smaller, if necessary, depending on δ.

For the term II we instead use (5.43):

$$\begin{aligned}\mathrm{II} &\leq 2\left(\frac{en}{n-k}\left(\frac{C_\delta}{\sqrt{n}}\right)^{\alpha/4}\right)^{2(n-k)}\\ &\leq 2\left(\frac{eC_\delta^{\alpha/4}n^{-\alpha/8}}{\varepsilon}\right)^{2(n-k)}.\end{aligned}$$

Hence, if

$$\varepsilon \geq 100C_\delta^{1/4}n^{-\alpha/8} =: K_\delta n^{-\alpha/8} \tag{5.46}$$

then

$$\begin{aligned}\mathrm{II} &\leq 2\exp\left(-2(n-k)\right)\\ &\leq 2\exp\left(-2K_\delta n^{7/8}\right).\end{aligned} \tag{5.47}$$

Together with (5.45) and (5.42) we conclude

$$\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1}) \ll_\delta \exp\left(-c_\delta n^{7/8}\right) \tag{5.48}$$

for $k \in [(1-\varepsilon_0)n, (1-\varepsilon)n]$, with ε satisfying (5.46). The result now follows from (5.4), summing our bounds (5.16), (5.37) and (5.48) for $\mathbf{P}(\mathcal{E}_k \setminus \mathcal{E}_{k-1})$ over $k \in [2, n-(1-\varepsilon)n]$. $\square$

Appendix A. Proofs of discrepancy properties

In this section we prove the discrepancy properties of Proposition 3.1 and Theorem 3.5, as well as a deterministic consequence of these properties, Lemma 5.3.

For distinct $i_1, i_2 \in [n]$, recall the statistic

$$h_{(i_1,i_2)}(M) = \left| \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \right|. \tag{A.1}$$

We will frequently use the fact that $h_{(i_1,i_2)}(M)$ determines the number of the other three possibilities for the 2×1 minor $M_{(i_1,i_2)\times j}$ through the row sums constraint:

$$\left| \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \right| = h_{(i_1,i_2)}(M), \tag{A.2}$$

$$\left| \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \right| = d - h_{(i_1,i_2)}(M), \text{ and} \tag{A.3}$$

$$\left| \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \right| = n - d - h_{(i_1,i_2)}(M). \tag{A.4}$$

We start with

Proposition 3.1 (Codegree discrepancy)**.** *For $\lambda \in (0,1)$ let*

$$\mathcal{B}_\lambda = \bigvee_{\substack{(i_1,i_2) \in [n]\times[n], \\ i_1 \neq i_2}} \left\{ \left| \frac{h_{(i_1,i_2)}(M)}{v_\delta n} - 1 \right| \geq \lambda \right\} \vee \left\{ \left| \frac{h_{(i_1,i_2)}(M^\mathsf{T})}{v_\delta n} - 1 \right| \geq \lambda \right\}$$

where $v_\delta = \delta(1-\delta)$. Then

$$\mathbf{P}(\mathcal{B}_\lambda) \ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta) n\right). \tag{3.3}$$

We will prove this using the *reflections* constructions from Section 3.2 (see Definitions 3.16 and 3.17) and Chatterjee's method of exchangeable pairs for concentration of measure (see Theorem A.3 below).

We begin with a crude version of Proposition 3.1.

Lemma A.1 (Crude codegree discrepancy)**.** *With probability $1 - O(n^2 \exp(-cv_\delta^3 n))$, for all $i_1 \neq i_2$ in $[n]$ we have*

$$h_{(i_1,i_2)}(M) \gg v_\delta^2 n,$$

where we recall $v_\delta := \delta(1-\delta)$.

Remark A.2. This lemma combines with the deterministic bound

$$h_{(i_1,i_2)}(M) \leq \min(d, n-d) = \min(\delta, 1-\delta)n + O(1) \tag{A.5}$$

(from (A.3) and (A.4)) to give that $h_{(i_1,i_2)}(M) \asymp_\delta n$ with overwhelming probability. This is actually sufficient for all of the applications of codegree discrepancy in sections 4 and 5, but we will need the more precise control provided by Proposition 3.1 to prove Theorem 3.5.

Proof. By row exchangeability and a union bound it suffices to show that $h_{(1,2)}(M) \gg v_\delta n$ except on an event of size $O(\exp(-cv_\delta^3 n))$. We denote $h(M) = h_{(1,2)}(M)$. We call a column index j "up" for M if

$$M_{(1,2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so that

$$h(M) = \sum_{j=1}^{n} 1_{\{j \text{ is up for } M\}}.$$

Draw M uniformly from $\mathcal{M}_d$. Let $m \geq 1$ to be chosen later, and draw the following random sequences independently of each other and of M:

(1) $\xi = (\xi_l)_{l=1}^m$ a sequences of i.i.d. $\mathrm{Ber}_\pm(1/2)$ random variables.
(2) $(J_l)_{l=1}^{2m}$ sampled uniformly without replacement from $[n]$. For each $l \in [m]$ we set $\mathbf{q}_l = (J_l^+, J_l^-) = (J_{2l-1}, J_{2l})$, and denote $\mathcal{Q} = (\mathbf{q}_l)_{l=1}^m$.

Now form $\widetilde{M} \stackrel{d}{=} M$ as in Corollary 3.19: for each $l \in [m]$, we perform a random reflection at $\mathbf{q}_l$ according to ξ_l. We will bound the probability that $h(M)$ is small by instead bounding the probability that $h(\widetilde{M})$ is small, using only the randomness of the $\mathbf{q}_l$ and ξ_l variables.

For $1 \leq l \leq m$, let $\mathcal{E}_l = \{\mathbf{q}_l \text{ is reflecting}\}$, and let $Z_\mathcal{Q} = \sum_{l=1}^n 1_{\mathcal{E}_l}$. Conditional on M and $\mathcal{Q}$, we have

$$h(\widetilde{M}) = h_0(M, \mathcal{Q}) + W \tag{A.6}$$

where W is a $\mathrm{Bin}(Z_\mathcal{Q}, 1/2)$ distributed random variable tallying the number of up indices j for $\widetilde{M}$ coming from reflections, and h_0 is a deterministic variable under this conditioning which tallies the rest of the up indices for $\widetilde{M}$. The indices contributing to h_0 are

(1) $j \in [n] \setminus \bigcup_{l=1}^m \{J_l^+, J_l^-\}$ such that j is up for M, and
(2) $j \in \bigcup_{l=1}^m \{J_l^+, J_l^-\}$ that are up for M but come from a non-reflecting pair.

We can ignore the contribution from the second category and simply bound $h_0 \geq h(M) - 2m$. Now we have

$$h(\widetilde{M}) \geq \max(h(M) - 2m, W).$$

Let

$$\mathcal{G} = \{h(M) \geq \varepsilon v_\delta n\}$$

for some $\varepsilon > 0$ to be chosen later, and take $m = \lfloor \frac{1}{4} \varepsilon v_\delta n \rfloor$. Then on $\mathcal{G}$ we have $h(\widetilde{M}) \geq \lfloor \frac{1}{2} \varepsilon v_\delta n \rfloor$, which is acceptable.

On $\mathcal{G}^c$, we show that W is large with overwhelming probability. First we argue that $Z_{\mathcal{Q}}$ is large. For each $l \in [m]$,

$$\begin{aligned}\mathbf{P}(\mathcal{E}_l|M) &\geq \mathbf{P}\left(M_{(1,2)\times(J_l^+,J_l^-)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \,\middle|\, M \right) \\ &\geq \frac{d-h(M)-2m}{n}\frac{n-d-h(M)-2m}{n} \\ &\geq (\delta - \varepsilon v_\delta - \frac{1}{2}\varepsilon v_\delta)(1-\delta-\varepsilon v_\delta - \frac{1}{2}\varepsilon v_\delta) \\ &\geq \frac{1}{4}\delta(1-\delta)\end{aligned}$$

for ε sufficiently small, where in the second line we have used (A.3) and (A.4). Applying case (2) of Corollary 3.22 with E the set of reflecting pairs $\mathbf{q}_l$ and $q_1 = \frac{1}{4}v_\delta$, we conclude that

$$\mathbf{P}(Z_{\mathcal{Q}} < cv_\delta m|M) \ll \exp(-cv_\delta^2 m)$$

for some $c > 0$ sufficiently small.

Now conditional on M such that $\mathcal{G}$ does not hold and conditional on $\mathcal{Q}$ such that $Z_{\mathcal{Q}} \geq cv_\delta m$, since W is $\mathrm{Bin}(Z_{\mathcal{Q}}, 1/2)$ distributed we have

$$W \geq c' v_\delta m$$

with probability $1 - \exp(-c' v_\delta^2 m)$, for some smaller constant $c' > 0$, by the lower bound on $Z_{\mathcal{Q}}$ and Hoeffding's inequality, say. Undoing the conditioning, we have that

$$h(M) \gg v_\delta m$$

with probability $1 - O(\exp(-c'' v_\delta^2 m))$. Substituting $m = \lfloor \frac{1}{4}\varepsilon v_\delta n \rfloor$ completes the proof. $\square$

We now upgrade Lemma A.1 to Proposition 3.1 using Chatterjee's method of exchangeable pairs for concentration of measure. The following is an abridged version of Theorem 1.5 from [14] suitable for our purposes:

Theorem A.3 (Chatterjee [14]). *Let $\mathcal{Z}$ be a separable metric space and suppose $(Z, \widetilde{Z})$ is an exchangeable pair of $\mathcal{Z}$-valued random variables. Suppose $f : \mathcal{Z} \to \mathbf{R}$ and $F : \mathcal{Z} \times \mathcal{Z} \to \mathbf{R}$ are square-integrable functions such that $F(Z, \widetilde{Z}) = -F(\widetilde{Z}, Z)$ a.s. and $\mathbf{E}(F(Z, \widetilde{Z})|Z) = f(Z)$ a.s.. Assume*

$$\mathbf{E}(e^{\theta f(Z)}|F(Z, \widetilde{Z})|) < \infty \tag{A.7}$$

for all θ. Let

$$\Delta(Z) := \frac{1}{2}\mathbf{E}\left(|(f(Z) - f(\widetilde{Z}))F(Z, \widetilde{Z})| \big| Z \right).$$

If there exists a constant C such that $\Delta(Z) \leq C$ a.s., then for any $t \geq 0$ we have

$$\mathbf{P}(|f(Z)| \geq t) \leq 2\exp(-t^2/2C). \tag{A.8}$$

Remark A.4. The qualitative integrability conditions on f and F will be satisfied automatically in our applications as we will only consider bounded (depending on n) functions on a finite set.

Now to prove Proposition 3.1, we first restrict to a "good" event on which all pairs of rows and columns of M are far from parallel. Let $\mathcal{G}$ denote the event that for all pairs of

row indices (i_1, i_2), $h_{(i_1,i_2)}(M)$ and $h_{(i_1,i_2)}(M^\mathsf{T})$ are at least $\varepsilon v_\delta^2 n$. By Lemma A.1 (applied to M and M^T), for ε sufficiently small we have

$$\mathbf{P}(\mathcal{G}^c) \ll n^2 \exp(-cv_\delta^3 n)$$

Hence it suffices to show

$$\mathbf{P}(\mathcal{B}_\lambda \wedge \mathcal{G}) \ll n^2 \exp(-c\lambda^2 v_\delta^2 n).$$

Furthermore, by row and column exchangeability and a union bound, it suffices to show

$$\mathbf{P}(|h(M) - v_\delta n| \geq \lambda v_\delta n) \ll \exp(-c\lambda^2 v_\delta^2 n)$$

where $h(M) := h_{(1,2)}(M) 1_{\mathcal{G}}$.

We define an exchangeable pair $(M, \widetilde{M})$ of random matrices from $\mathcal{M}_d$ as follows. We draw M uniformly at random, and additionally draw

(1) J_1, J_2 uniformly from $[n]$, independently of each other and of M, and
(2) a random permutation σ uniformly from $\mathrm{Sym}(n-2)$, independent of M, J_1, J_2.

First we permute the rows $R_3, \dots, R_n$ of M according to σ to form a matrix M^σ. Then we form $\widetilde{M}$ by performing a reflection at (J_1, J_2) on M^σ. $(M, \widetilde{M})$ is an exchangeable pair of random matrices uniformly distributed on $\mathcal{M}_d$ by row-exchangeability and the fact from (3.31) that the reflection transformation is an involution.

Note that if (J_1, J_2) is reflecting then we have that $M_{(1,2,I^*)\times(J_1,J_2)}$ is equal to either

$$\mathbf{A} := \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{B} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and a reflection at (J_1, J_2) sends $\mathbf{A}$ to $\mathbf{B}$ and $\mathbf{B}$ to $\mathbf{A}$. Let $\mathcal{E}_A$ denote the event that (J_1, J_2) is reflecting and $M_{(1,2,I^*)\times(J_1,J_2)}$ is equal to $\mathbf{A}$, and let $\mathcal{E}_B$ denote the event that (J_1, J_2) is reflecting and that $M_{(1,2,I^*)\times(J_1,J_2)} = \mathbf{B}$, so that $\mathcal{E} := \mathcal{E}_A \vee \mathcal{E}_B$ is the event that (J_1, J_2) is reflecting. Since a reflection only occurs on the event $\mathcal{E}$, we have $\widetilde{M} = M^\sigma$ on $(\mathcal{E}_A \vee \mathcal{E}_B)^c$.

Define $F(M, \widetilde{M}) = h(M) - h(\widetilde{M})$. Letting

$$\mathcal{E}_j = \left\{ M_{(1,2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \wedge \mathcal{G}, \quad \widetilde{\mathcal{E}}_j = \left\{ \widetilde{M}_{(1,2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \wedge \mathcal{G}$$

we have $h(M) = \sum_{j=1}^n 1_{\mathcal{E}_j}$ and $h(\widetilde{M}) = \sum_{j=1}^n 1_{\widetilde{\mathcal{E}}_j}$. Since M and $\widetilde{M}$ only differ on the columns J_1, J_2, and only if $\mathcal{E}_A$ or $\mathcal{E}_B$ holds, we have

$$\begin{aligned} F(M, \widetilde{M}) &= (1_{\mathcal{E}_{J_1}} - 1_{\widetilde{\mathcal{E}}_{J_1}} + 1_{\mathcal{E}_{J_2}} - 1_{\widetilde{\mathcal{E}}_{J_2}}) 1_{\mathcal{E}_A \vee \mathcal{E}_B} \\ &= (1_{\mathcal{E}_{J_1}} - 1_{\widetilde{\mathcal{E}}_{J_1}}) 1_{\mathcal{E}_A \vee \mathcal{E}_B} \\ &= 1_{\mathcal{E}_B \wedge \mathcal{G}} - 1_{\mathcal{E}_A \wedge \mathcal{G}}. \end{aligned}$$

Now we compute $f(M) := \mathbf{E}(F(M,\widetilde{M})|M)$ (restricting M to the event $\mathcal{G}$). On the one hand, we have

$$\begin{aligned}\mathbf{E}(1_{\mathcal{E}_A}|M) &= \mathbf{P}(\mathcal{E}_A|M)\\ &= \mathbf{P}\left(\mathcal{E}_A \wedge \left\{M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}\right\} \,\middle|\, M\right)\\ &= \mathbf{P}\left(M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} \,\middle|\, M\right)\\ &= \frac{d-h(M)}{n}\,\frac{n-d-h(M)}{n} \end{aligned}\tag{A.9}$$

where we have used the fact that (J_1, J_2) is automatically reflecting if $M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$, and the last line follows from (A.3) and (A.4).

For the other term, we have

$$\begin{aligned}\mathbf{E}(1_{\mathcal{E}_B}|M) &= \mathbf{P}(\mathcal{E}_B|M)\\ &= \mathbf{P}\left(\mathcal{E}_B \wedge \left\{M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right\} \,\middle|\, M\right)\\ &= \mathbf{P}_{\sigma,J_1,J_2}\left(\mathcal{E} \,\middle|\, M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\mathbf{P}\left(M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \,\middle|\, M\right)\\ &= \frac{h(M)^2}{n^2}\left(1 - \mathbf{P}_{\sigma,J_1,J_2}\left(\mathcal{E}^c \,\middle|\, M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\right) \end{aligned}\tag{A.10}$$

where in the last line we have used (A.1) and (A.2).

The random permutation σ was included to handle this last term:

Claim A.5. $\mathbf{P}_{\sigma,J_1,J_2}\left(\mathcal{E}^c \mid M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) \ll \frac{1}{v_\delta^2 n}$.

Combining (A.9), (A.10) and Claim A.5 we have

$$\begin{aligned}f(M) &= \frac{1}{n}(h(M) - v_\delta n) + O\left(\frac{h(M)^2}{v_\delta^2 n^3}\right)\\ &= \frac{1}{n}(h(M) - v_\delta n) + O(1/n)\end{aligned}$$

where we have used (A.5) in the second line. (In particular it follows from the anti-symmetry of F that $\mathbf{E}\, f(M) = 0$ and hence $\mathbf{E}\, h(M) = v_\delta n + O(1)$.)

Now

$$\begin{aligned}\Delta(M) &:= \frac{1}{2}\mathbf{E}\Big(|(f(M) - f(\widetilde{M}))F(M,\widetilde{M})|\big|M\Big)\\ &= \frac{1}{2n}\mathbf{E}\Big(|h(M) - h(\widetilde{M})|^2 + O(1)\big|M\Big)\\ &= O(1/n)\end{aligned}$$

where we have used that the reflection producing $\widetilde{M}$ from M can change $h(M)$ by at most 1. We conclude from Theorem A.3 that for any $\lambda > 0$,

$$\mathbf{P}(|h(M) - v_\delta n| \geq \lambda v_\delta n) \leq 2\exp(-c\lambda^2 v_\delta^2 n).$$

This concludes the proof of Proposition 3.1, on Claim A.5.

Proof of Claim A.5. By column exchangeability we have

$$\mathbf{P}_{\sigma,J_1,J_2}\left(\mathcal{E}^c \,\middle|\, M_{(1,2)\times(J_1,J_2)} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) = \mathbf{P}_\sigma\big((1,2) \text{ not reflecting}|M_{(1,2)\times(1,2)} = \mathbf{I}_2\big).$$

Now recall from Definition 3.16 that given $M_{(1,2)\times(1,2)} = \mathbf{I}_2$, $(1,2)$ is not reflecting for M^σ if and only if there is no row index I such that

$$\sum_{i=2}^{I} M(i,1) = \sum_{i=2}^{I} M(i,2).$$

In terms of the associated walk

$$\begin{aligned} w(I) &= \sum_{i=1}^{I} 1_{\{M_{i\times(j_1,j_2)}=(0\quad 1)\}} - 1_{\{M_{i\times(j_1,j_2)}=(1\quad 0)\}} \\ &=: w_+(I) - w_-(I) \end{aligned}$$

this is the event that $w(1) = -1$, $w(2) = 0$, and there is not $I \geq 3$ such that $w(I) = -1$. By the column sums constraint we have

$$w_+(n) = w_-(n) = h_{(1,2)}(M^\mathsf{T}) =: h$$

i.e. the walk takes $2h$ steps and ends at 0. Since we are on $\mathcal{G}$ we have $h \gg v_\delta^2 n$. The event that $(1,2)$ is not reflecting for M^σ is the event that $w(I) \geq 0$ for $3 \leq I \leq n$, i.e. that it is "non-crossing" for this range of I. The number of non-crossing walks with $2h-2$ steps is given by the Catalan number $\frac{1}{h}\binom{2(h-1)}{h-1}$, which is a fraction $1/h$ of the total number of walks. Since under the randomness of σ each walk is equally likely, we have

$$\begin{aligned} \mathbf{P}_\sigma\big((1,2) \text{ not reflecting}|M_{(1,2)\times(1,2)} = \mathbf{I}_2\big) &= \mathbf{P}_\sigma(w(I) \text{ non-crossing}) \\ &= 1/h \\ &\ll 1/v_\delta^2 n. \end{aligned}$$

□

Now we turn to the edge discrepancy property. The proof again employs Theorem A.3, leveraging the codegree discrepancy property to locate switchable 2×2 minors.

Theorem 3.5 (Edge discrepancy)**.** *For $K > 0$, define the "coarse-scale" family of pairs of subsets of $[n]$*

$$\mathcal{F}_\mathfrak{c}(K) = \{(S,T) : |S||T| \geq Kn\} \tag{3.12}$$

and the larger "fine-scale" family

$$\mathcal{F}_\mathfrak{f}(K) = \left\{(S,T) : |S||T| \geq Kn\, G\left(\frac{\max(S,T)}{n}\right)\right\} \tag{3.13}$$

where $G(x) := -x\log x$. For $\lambda \in (0,1)$ and $\mathcal{F}$ a family of pairs of subsets of $[n]$, define the bad event

$$\mathcal{B}_{\lambda,\mathcal{F}} := \left\{\exists\,(S,T) \in \mathcal{F} :\ \left|\frac{e_{S,T}(M)}{\delta|S||T|} - 1\right| \geq \lambda\right\}. \tag{3.14}$$

For any $\lambda \in (0,1)$ there exists $K_{\lambda,\delta} > 0$ depending only on λ, δ such that

$$\mathbf{P}\Big(\mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{c}}(K_{\lambda,\delta})}\Big) \ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta) n\right) \tag{3.15}$$

and

$$\mathbf{P}\Big(\mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{f}}(K_{\lambda,\delta})}\Big) \ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta) n\right) + \exp\left(-c \log^2 n\right). \tag{3.16}$$

In fact we can take $K_{\lambda,\delta} = \frac{C}{v_\delta^3 \lambda^2 (1-\lambda)^2}$ for a sufficiently large constant C.

The core of the proof is the following tail estimate for the number of edges passing from S to T for fixed sets of vertices $S, T \subset [n]$.

Lemma A.6 (Concentration of edge counts)**.** *Let $S, T \subset [n]$ be subsets of row indices and column indices, respectively. Let $\lambda \in (0,1)$, and let $\mathcal{B}_\lambda$ denote the event from Proposition 3.1 that the codegree discrepancy property does not hold. Then for any $N \geq 0$,*

$$\mathbf{P}\big(\mathcal{B}_\lambda^c \wedge \big|e_{S,T}(M) - \delta|S||T|\big| \geq N\big) \leq 2\exp\left(-cv_\delta \frac{(1-\lambda)^2 n^2 N^2}{|S|(n-|S|)|T|(n-|T|)}\right). \tag{A.11}$$

Proof. Fix $S, T \subset [n]$ and put $|S| = s, |T| = t$. Draw M uniformly at random from $\mathcal{M}_d$. Draw $I_1 \in S$, $I_2 \in [n] \setminus S$, $J_1 \in T$, $J_2 \in [n] \setminus T$ independently of each other and M, and uniformly from their respective ranges. Conditional on M, form $\widetilde{M}$ from M by performing a switching at $(I_1, I_2) \times (J_1, J_2)$. $(M, \widetilde{M})$ is an exchangeable pair of random matrices uniformly distributed on $\mathcal{M}_d$ by Lemma 3.11.

Let A denote the randomly sampled 2×2 minor $M_{(I_1,I_2)\times(J_1,J_2)}$, and let $\mathcal{E} = \{A = \mathbf{I}_2\} \vee \{A = \mathbf{J}_2\}$ be the event that we sample a switchable minor. Then we have

$$\begin{aligned} F(M, \widetilde{M}) &:= (e_{S,T}(M) - e_{S,T}(\widetilde{M})) 1_{\mathcal{B}_\lambda^c} \\ &= \Big(1_{\{M(I_1,J_1)=1\}\wedge\mathcal{B}_\lambda^c} - 1_{\{\widetilde{M}(I_1,J_1)=1\}\wedge\mathcal{B}_\lambda^c}\Big) 1_{\mathcal{E}}. \end{aligned}$$

(Note that $\mathcal{B}_\lambda$ holds for M if and only if it holds for $\widetilde{M}$, so that $F(M, \widetilde{M})$ is anti-symmetric.) Conditional on M such that $\mathcal{B}_\lambda$ does not hold,

$$\begin{aligned} \mathbf{E}(F(M, \widetilde{M})|M) &= \mathbf{P}(\mathcal{E} \wedge \{M(I_1, J_1) = 1\} |M) - \mathbf{P}\Big(\mathcal{E} \wedge \Big\{\widetilde{M}(I_1, J_1) = 1\Big\} |M\Big) \\ &= \mathbf{P}(A = \mathbf{I}_2 | M) - \mathbf{P}(A = \mathbf{J}_2 | M). \end{aligned}$$

For $i \in [n]$, let

$$\begin{aligned} T_i(M) &:= \{j \in T : M(i,j) = 1\} \qquad \text{(A.12)} \\ &= \sum_{j \in T} M(i,j), \qquad \text{(A.13)} \end{aligned}$$

and for $i_1 \in S$, $i_2 \in [n] \setminus S$ and $a, b \in \{0,1\}$ let

$$\hat{T}^{ab}_{(i_1,i_2)}(M) := \left\{ j \in T : M_{(i_1,i_2)\times j} = \begin{pmatrix} a \\ b \end{pmatrix} \right\} \tag{A.14}$$

$$\check{T}^{ab}_{(i_1,i_2)}(M) := \left\{ j \in [n] \setminus T : M_{(i_1,i_2)\times j} = \begin{pmatrix} a \\ b \end{pmatrix} \right\}. \tag{A.15}$$

Recalling

$$h_{(i_1,i_2)}(M) := \left| \left\{ j \in [n] : M_{(i_1,i_2)\times j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \right|$$

the following four identities are consequences of the row sums constraint:

$$h_{(i_1,i_2)}(M) = |\hat{T}^{10}_{(i_1,i_2)}(M)| + |\check{T}^{10}_{(i_1,i_2)}(M)| \tag{A.16}$$
$$= |\hat{T}^{01}_{(i_1,i_2)}(M)| + |\check{T}^{01}_{(i_1,i_2)}(M)|, \tag{A.17}$$
$$|T_{i_1}| = |\hat{T}^{11}_{(I_1,I_2)}(M)| + |\hat{T}^{10}_{(i_1,i_2)}(M)|, \tag{A.18}$$
$$|T_{i_2}| = |\hat{T}^{11}_{(I_1,I_2)}(M)| + |\hat{T}^{01}_{(i_1,i_2)}(M)|. \tag{A.19}$$

In terms of these quantities we express

$$\mathbf{P}(A = \mathbf{I}_2 | M, I_1, I_2) = \frac{|\hat{T}^{10}_{(I_1,I_2)}(M)||\check{T}^{01}_{(I_1,I_2)}(M)|}{t(n-t)}$$

and

$$\mathbf{P}(A = \mathbf{J}_2 | M, I_1, I_2) = \frac{|\hat{T}^{01}_{(I_1,I_2)}(M)||\check{T}^{10}_{(I_1,I_2)}(M)|}{t(n-t)}.$$

Suppressing dependences on M, I_1, I_2, and assuming M is such that $\mathcal{B}_\lambda$ does not hold, we have

$$\begin{aligned}
\mathbf{E}(F(M, \widetilde{M})|M, I_1, I_2) &= \frac{1}{t(n-t)}(|\hat{T}^{10}||\check{T}^{01}| - |\hat{T}^{01}||\check{T}^{10}|) \\
&= \frac{1}{t(n-t)}(|\hat{T}^{10}|(h - |\hat{T}^{01}|) - |\hat{T}^{01}|(h - |\hat{T}^{10}|)) \\
&= \frac{1}{t(n-t)} h(|\hat{T}^{10}| - |\hat{T}^{01}|) \\
&= \frac{h(|T_{I_1}| - |T_{I_2}|)}{t(n-t)}
\end{aligned}$$

where in the second line we have used (A.16) and (A.17), and in the fourth line we have used (A.18) and (A.19). In full regalia,

$$\mathbf{E}(F(M, \widetilde{M})|M, I_1, I_2) = \frac{1}{t(n-t)} h_{(I_1,I_2)}(M)(|T_{I_1}(M)| - |T_{I_2}(M)|) 1_{\mathcal{B}^c_\lambda}.$$

Note that

$$\begin{aligned}
\sum_{i_1 \in S} |T_{i_1}(M)| &= \sum_{i_1 \in S} \sum_{j_1 \in T} M(i_1, j_1) \\
&= e_{S,T}(M)
\end{aligned}$$

and similarly

$$\begin{aligned}
\sum_{i_2 \in [n] \setminus S} |T_{i_2}(M)| &= e_{[n]\setminus S, T}(M) \\
&= td - e_{S,T}(M)
\end{aligned}$$

where the last line follows from the column-sums constraint. It follows that

$$\begin{aligned}
f(M) &:= \mathbf{E}(F(M,\widetilde{M})|M) \\
&= \frac{h_{(I_1,I_2)}(M)}{s(n-s)t(n-t)} 1_{\mathcal{B}_\lambda^c} \sum_{i_1\in S, i_2\in[n]\setminus S} |T_{I_1}(M)| - |T_{I_2}(M)| \\
&= \frac{h_{(I_1,I_2)}(M)}{s(n-s)t(n-t)} \big((n-s)e_{S,T}(M) - s(td - e_{S,T}(M))\big) 1_{\mathcal{B}_\lambda^c} \\
&= \frac{nh_{(I_1,I_2)}(M)}{s(n-s)t(n-t)} (e_{S,T}(M) - \delta st) 1_{\mathcal{B}_\lambda^c}.
\end{aligned}$$

In particular, by the restriction to $\mathcal{B}_\lambda^c$,

$$|f(M)| \in (1-\lambda, 1+\lambda) \frac{v_\delta n^2}{s(n-s)t(n-t)} |e_{S,T}(M) - \delta st| 1_{\mathcal{B}_\lambda^c}. \tag{A.20}$$

Now we have the bound

$$\begin{aligned}
\Delta(M) &:= \frac{1}{2}\mathbf{E}\Big(\big|(f(M) - f(\widetilde{M}))F(M,\widetilde{M})\big|\big|M\Big) \\
&= \frac{1}{2}\frac{nh_{(I_1,I_2)}(M)}{s(n-s)t(n-t)} \mathbf{E}\Big(|e_{S,T}(M) - e_{S,T}(\widetilde{M})|^2\big|M\Big) 1_{\mathcal{B}_\lambda^c} \\
&\le \frac{1+\lambda}{2} \frac{v_\delta n^2}{s(n-s)t(n-t)}
\end{aligned}$$

where we have used that $e_{S,T}(M)$ and $e_{S,T}(\widetilde{M})$ differ by at most 1, and that M is in the complement of $\mathcal{B}_\lambda$. By Theorem A.3 and the left endpoint of (A.20) we have

$$\begin{aligned}
\mathbf{P}\big(\mathcal{B}_\lambda^c \wedge \big|e_{S,T}(M) - \delta|S||T|\big| \ge N\big) &\le \mathbf{P}\bigg(|f(M)| \ge \frac{(1-\lambda)v_\delta n^2}{s(n-s)t(n-t)} N\bigg) \\
&\le 2\exp\left(-cv_\delta \frac{(1-\lambda)^2 n^2 N^2}{s(n-s)t(n-t)}\right)
\end{aligned}$$

which proves (A.11). □

Now we conclude the proof of Theorem 3.5. Let $\mathcal{B}_\lambda$ be as in Proposition 3.1, and recall

$$\mathcal{B}_{\lambda,\mathcal{F}_\mathfrak{c}(K)} := \left\{\exists S,T\subset[n]:\ \ |S||T| \ge Kn,\ \ \left|\frac{e_{S,T}(M)}{\delta|S||T|} - 1\right| \ge \lambda\right\}.$$

By a union bound, Proposition 3.1 and Lemma A.6,

$$\begin{aligned}
\mathbf{P}\big(\mathcal{B}_{\lambda,\mathcal{F}_\mathfrak{c}(K)}\big) &= \mathbf{P}\big(\mathcal{B}_\lambda \wedge \mathcal{B}_{\lambda,\mathcal{F}_\mathfrak{c}(K)}\big) + \mathbf{P}\big(\mathcal{B}_\lambda^c \wedge \mathcal{B}_{\lambda,\mathcal{F}_\mathfrak{c}(K)}\big) \\
&\le \mathbf{P}(\mathcal{B}_\lambda) + \sum_{S,T\subset[n]:|S||T|\ge Kn} \mathbf{P}\left(\mathcal{B}_\lambda^c \wedge \left\{\left|\frac{e_{S,T}(M)}{\delta|S||T|} - 1\right| \ge \lambda\right\}\right) \\
&\le \mathbf{P}(\mathcal{B}_\lambda) + 2^{2n} 2\exp\left(-cv_\delta^3 \frac{\lambda^2(1-\lambda)^2 Kn^3}{(n-|S|)(n-|T|)}\right) \\
&\ll \mathbf{P}(\mathcal{B}_\lambda) + \exp(-cn) \\
&\ll n^2 \exp\big(-cv_\delta^2 \min(\lambda, v_\delta) n\big)
\end{aligned}$$

where we have taken

$$K = K_{\lambda,\delta} := \frac{C}{v_\delta^3 \lambda^2 (1-\lambda)^2} \tag{A.21}$$

for a sufficiently large absolute constant C . This proves (3.15).

For the bound (3.16), recall that with $G(x) := -x \log x$,

$$\begin{aligned} \mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{l}}(K)} &:= \left\{ \exists S,T \subset [n] : \ |S||T| \geq Kn\, G\left(\frac{\max(S,T)}{n}\right), \ \left| \frac{e_{S,T}(M)}{\delta|S||T|} - 1 \right| \geq \lambda \right\} \\ &= \left\{ \exists S,T \subset [n] : \ \min(|S|,|T|) \geq K \log \frac{n}{\max(|S|,|T|)}, \ \left| \frac{e_{S,T}(M)}{\delta|S||T|} - 1 \right| \geq \lambda \right\}. \end{aligned}$$

Let

$$E = \left\{ (s,t) \in [n] \times [n] : \min(s,t) \geq K \log \frac{n}{\max(s,t)} \right\}.$$

Then again by a union bound, Proposition 3.1 and Lemma A.6 we have

$$\begin{aligned} \mathbf{P}\big(\mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{c}}(K)}\big) &\leq \mathbf{P}\big(\mathcal{B}_\lambda \wedge \mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{c}}(K)}\big) + \mathbf{P}\big(\mathcal{B}_\lambda^c \wedge \mathcal{B}_{\lambda,\mathcal{F}_{\mathfrak{c}}(K)}\big) \\ &\leq \mathbf{P}(\mathcal{B}_\lambda) + \sum_{(S,T)\in\mathcal{F}_{\mathfrak{f}}(K)} \mathbf{P}\left(\mathcal{B}_\lambda^c \wedge \left\{ \left| \frac{e_{S,T}(M)}{\delta|S||T|} - 1 \right| \geq \lambda \right\} \right) \\ &\ll \mathbf{P}(\mathcal{B}_\lambda) + \sum_{(s,t)\in E} \binom{n}{s}\binom{n}{t} \exp\left(-cv_\delta^3 \lambda^2 (1-\lambda)^2 st\right) \\ &\ll \mathbf{P}(\mathcal{B}_\lambda) + \sum_{(s,t)\in E} \exp\left(s\left(1 + \log\frac{n}{s}\right) + t\left(1 + \log\frac{n}{t}\right) - cv_\delta^3\lambda^2(1-\lambda)^2 st \right) \\ &\ll \mathbf{P}(\mathcal{B}_\lambda) + \sum_{(s,t)\in E} \exp\left(C\max(s,t)\left(\log\frac{n}{\max(s,t)} - cv_\delta^3\lambda^2(1-\lambda)^2 \min(s,t) \right) \right) \\ &\ll \mathbf{P}(\mathcal{B}_\lambda) + \sum_{(s,t)\in E} \exp\left(-C(Kcv_\delta^3\lambda^2(1-\lambda)^2 - 1)\max(s,t)\log\frac{n}{\max(s,t)} \right) \\ &\ll n^2 \exp\left(-cv_\delta^2 \min(\lambda, v_\delta) n\right) + \exp\left(-c\log^2 n\right) \end{aligned}$$

for $K = K_{\lambda,\delta}$ as in (A.21). □

The following deterministic consequence of the codegree and edge discrepancy properties is needed in the proof of Proposition 2.6.

Lemma 5.3. *Let $A \in \mathcal{M}_d$ and suppose that for some $\lambda > 0$ sufficiently small depending on δ and $K > 0$, A has codegree and edge discrepancy properties: for all distinct $i_1, i_2 \in [n]$,*

$$h_{(i_1,i_2)\times[n]}(A) \quad \text{and} \quad h_{(i_1,i_2)\times[n]}(A^\mathsf{T}) \in (1-\lambda, 1+\lambda)v_\delta n \tag{5.19}$$

and for all pairs $S_0, T_0 \subset [n]$ such that $|S_0||T_0| \geq Kn$ we have

$$e_{S_0,T_0}(A) \in (1-\lambda, 1+\lambda)\delta|S_0||T_0|. \tag{5.20}$$

Then for any fixed $S, T \subset [n]$ with $|S||T| \geq \frac{10K}{\lambda}n$, there exist disjoint subsets $S^+, S^- \subset S$ with

$$|S^+| = |S^-| \gg \lambda^2\delta|S|$$

such that for each $i^+ \in S^+$, we have

$$h_{(i^+,i^-)\times T}(A) = \big(1+O(\lambda)\big)v_\delta|T| \tag{5.21}$$

$$h_{(i^-,i^+)\times T}(A) = \big(1+O(\lambda)\big)v_\delta|T| \tag{5.22}$$

$$h_{(i^+,i^-)\times [n]\setminus T}(A) = \big(1+O(\lambda)\big)v_\delta(n-|T|) \tag{5.23}$$

$$h_{(i^-,i^+)\times [n]\setminus T}(A) = \big(1+O(\lambda)\big)v_\delta(n-|T|) \tag{5.24}$$

for all but $O\big(\frac{Kn}{v_\delta|T|}\big)$ elements i^- of S^-.

Proof. Fix S, T as above. For each $i \in S$, let

$$T(i) = \{j \in T : A(i,j) = 1\}.$$

First we locate a subset S'' of S such that for each $i \in S''$,

$$|T(i)| = \sum_{j\in T} A(i,j)$$

is close to $\delta|T|$. Since $|S||T| \geq Kn$,

$$\sum_{i\in S} |T(i)| = e_{S,T}(A) < (1+\lambda)\delta|S||T|.$$

It follows (from Markov's inequality) that

$$|T(i)| \leq (1+2\lambda)\delta|T|$$

for at least $\frac{\lambda}{1+2\lambda}|S|$ values of $i \in S$. Let $S' \subset S$ be a set of $\geq \frac{\lambda}{1+2\lambda}|S|$ such i.

Now we pass to a further subset S'' on which $|T(i)|$ is not too small. Suppose that $|T(i)| \leq (1-2\lambda)\delta|T|$ for $\eta|S'|$ values of $i \in S'$. Then since

$$\begin{aligned}|S'||T| &\geq \frac{\lambda}{1+2\lambda}|S||T| \\ &\geq \frac{10}{1+2\lambda}Kn \\ &\geq Kn\end{aligned}$$

we can apply (5.20) and the upper bound $|T(i)| \leq |T|$ to get

$$\begin{aligned}\eta|S'|(1-2\lambda)\delta|T| + |T|(1-\eta)|S'| &\geq e_{S',T}(A) \\ &> (1-\lambda)\delta|S'||T|.\end{aligned}$$

Rearranging gives

$$\eta \leq \frac{1-\delta(1-\lambda)}{1-\delta(1-2\lambda)}.$$

Hence there is a subset $S'' \subset S$ with

$$\begin{aligned}|S''| &\geq (1-\eta)|S'| \\ &\geq \frac{\lambda\delta}{1-\delta(1-2\lambda)}\frac{\lambda}{1+2\lambda}|S| \\ &\gg \lambda^2\delta|S|\end{aligned}$$

for λ sufficiently small, such that for all $i \in S''$,

$$|T(i)| = (1+O(\lambda))\delta|T|. \tag{A.22}$$

Now let $S^+ \subset S''$ with $|S''| \ll |S^+| \leq |S|/2$ (which may be obtained from S'' by deleting elements until it has fewer than $|S|/2$ elements), and let S^- be an arbitrary subset of $S \setminus S^+$ with $|S^-| = |S^+|$.

Let $a, b \in (0,1)$ to be chosen later. For each $i^+ \in S^+$, define the "bad" subset of S^-

$$\begin{aligned} S(i^+) &= \big\{ i^- \in S^- : h_{(i^+,i^-)\times T}(A) \leq a|T(i^+)| \big\} \\ &\quad \cup \big\{ i^- \in S^- : h_{(i^+,i^-)\times T}(A) \geq (1-b)|T(i^+)| \big\} \\ &\quad \cup \big\{ i^- \in S^- : h_{(i^-,i^+)\times T}(A) \leq b(|T| - |T(i^+)|) \big\} \\ &\quad \cup \big\{ i^- \in S^- : h_{(i^-,i^+)\times T}(A) \geq (1-a)(|T| - |T(i^+)|) \big\} \\ &=: S_1(i^+) \cup S_2(i^+) \cup S_3(i^+) \cup S_4(i^+). \end{aligned} \tag{A.23}$$

It remains to bound $|S(i^+)|$.

Suppose $|S(i^+)| = m$ for some $i^+ \in S^+$, and let $\alpha \in (0, \frac{1}{2})$ to be optimized later. Then by pigeonholing, at least one of the following holds:

Case 1: $|S_1(i^+)| \geq \alpha m$.
Case 2: $|S_2(i^+)| \geq \alpha m$.
Case 3: $|S_3(i^+)| \geq (\frac{1}{2} - \alpha) m$.
Case 4: $|S_4(i^+)| \geq (\frac{1}{2} - \alpha) m$.

We show that if m is too large, each of these cases implies the existence of a large minor of A that is either more dense or more sparse than the edge discrepancy property allows.

Assume Case 1 holds, and suppose that

$$m \geq \frac{Kn}{\alpha|T(i^+)|}.$$

Then

$$|S_1(i^+)||T(i^+)| \geq \alpha m |T(i^+)| \geq Kn$$

so that by the bound (5.20) on edge discrepancy we have

$$e_{S_1(i^+),T(i^+)}(A) < (1+\lambda)\delta |S_1(i^+)||T(i^+)|. \tag{A.24}$$

On the other hand, noting that

$$h_{(i^+,i^-)\times T}(A) = |\{ j \in T(i^+) : A(i^-, j) = 0\}|$$

we have from the definition of $S_1(i^+)$ that

$$|\{ j \in T(i^+) : A(i^-, j) = 1\}| > (1-a)|T(i^+)|$$

for every $i^- \in S_1(i^+)$. Summing over $i^- \in S_1(i^+)$ we have

$$\begin{aligned} (1-a)|S_1(i^+)||T(i^+)| &< \sum_{i \in S_1(i^+)} \sum_{j \in T(i^+)} A(i,j) \\ &= e_{S_1(i^+),T(i^+)}(A) \end{aligned}$$

which with (A.24) implies

$$a > 1 - (1+\lambda)\delta.$$

Hence, taking $a = 1 - (1+\lambda)\delta$, we must have

$$m \leq \frac{Kn}{\alpha|T(i^+)|} \tag{A.25}$$

in Case 1.

A similar argument shows that in Case 2, taking $b = (1-\lambda)\delta$ ensures by the edge discrepancy property that (A.25) holds, and in Cases 3 and 4 these values of a, b similarly imply

$$|S(i^+)| = m \leq \frac{Kn}{(\frac{1}{2}-\alpha)(|T| - |T(i^+)|)}.$$

Hence,

$$|S(i^+)| \leq Kn \max\left(\frac{1}{\alpha|T(i)|}, \frac{1}{(\frac{1}{2}-\alpha)(|T| - |T(i)|)}\right).$$

Optimizing α gives

$$|S(i^+)| \leq \frac{2K}{\frac{|T(i^+)|}{|T|}\left(1 - \frac{|T(i^+)|}{|T|}\right)} \frac{n}{|T|}$$

Inserting the estimate (A.22) we have

$$|S(i^+)| \ll \frac{Kn}{v_\delta |T|} \tag{A.26}$$

(using that $\delta \leq 1/2$).

Now for any i^- outside $S(i^+)$, using (A.22) and our values for a, b we have

$$\begin{aligned} h_{(i^+,i^-)\times T}(A) &\geq a|T(i)| \\ &\geq (1-(1+\lambda)\delta)\delta(1-2\lambda)|T| \\ &= (1+O(\lambda))v_\delta|T|, \end{aligned}$$

and

$$\begin{aligned} h_{(i^+,i^-)\times T}(A) &\leq (1-b)|T(i)| \\ &\leq (1-(1-\lambda)\delta)\delta(1+2\lambda)|T| \\ &= (1+O(\lambda))v_\delta|T|. \end{aligned}$$

Using that

$$h_{(i^+,i^-)\times T}(A) + h_{(i^+,i^-)\times [n]\setminus T}(A) = h_{(i^+,i^-)}(A) = (1+O(\lambda))v_\delta n$$

(from (5.19)), we get

$$\begin{aligned} h_{(i^+,i^-)\times [n]\setminus T}(A) &= h_{(i^+,i^-)}(A) - h_{(i^+,i^-)\times T}(A) \\ &= (1+O(\lambda))v_\delta(n - |T|). \end{aligned}$$

We similarly conclude

$$\begin{aligned} h_{(i^-,i^+)\times T}(A) &= (1+O(\lambda))v_\delta|T| \\ h_{(i^-,i^+)\times [n]\setminus T}(A) &= (1+O(\lambda))v_\delta(n-|T|) \end{aligned}$$

for all $i^+ \in S^+$ and $i^- \in S^- \setminus S(i^+)$. $\square$

References

[1] N. Alon and J. Spencer, The probabilistic method. Third edition. With an appendix on the life and work of Paul Erdős. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2008.

[2] R. Adamczak and D. Chafaï, Circular law for random matrices with unconditional log-concave distribution. *Preprint available at arXiv:1303.5838.*

[3] R. Adamczak, D. Chafaï and P. Wolff, Circular law for random matrices with exchangeable entries. *Preprint available at arXiv:1402.3660.*

[4] N. Alon, M. Krivelevich and B. Sudakov, List coloring of random and pseudo-random graphs. *Combinatorica* **19** (1999), no. 4, 453–472.

[5] R. Arratia and S. Tavare, Independent process approximations for random combinatorial structures. *Adv. Math.* (1994) **104**, 90–154.

[6] A. Basak and A. Dembo, Limiting spectral distribution of sum of unitary and orthogonal matrices. *Electron. Commun. Probab.* **18** (2013), no. 69, 19 pp.

[7] C. Bordenave, P. Caputo and D. Chafaï, Circular law theorem for random Markov matrices. *Probab. Theory Related Fields* **152** (2012), no. 3-4, 751–779.

[8] C. Bordenave, P. Caputo and D. Chafaï, Spectrum of Markov generators on sparse random graphs. *To appear in Comm. Pure Appl. Math.. Available at arXiv:1202.0644.*

[9] C. Bordenave and D. Chafai, Around the circular law. *Probability Surveys* **9** (2012) 1–89.

[10] C. Bordenave and M. Lelarge, Resolvent of large random graphs. *Random Structures Algorithms* **37** (2010), no. 3, 332–352.

[11] J. Bourgain, P. Wood and V. Vu, On the singularity probability of random discrete matrices. *J. Funct. Anal.* **258** (2010), no. 2, p. 559–603.

[12] W. Bryc, A. Dembo and T. Jiang, Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34** (2006), no. 1, 1–38.

[13] E. R. Canfield and B. D. McKay, Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums. *Electron. J. Combin.* **12** (2005), *Research Paper 29, 31 pp. (electronic).*

[14] S. Chatterjee, Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321.

[15] S. Chatterjee, P. Diaconis and A. Sly, On properties of random doubly stochastic matrices, to appear in Annales de l'Institute Henri Poincaré. *To appear in Ann. de l'Inst. Henri Poincaré. Available at arXiv:math/1010.6136.*

[16] I. Dumitriu and T. Johnson, The Marcenko-Pastur law for sparse random bipartite biregular graphs. *Available at arXiv:math/1304.4907.*

[17] I. Dumitriu, T. Johnson, S. Pal and E. Paquette, Functional limit theorems for random regular graphs. *Probab. Theory Related Fields* **156** (2013), no. 3-4, 921–975.

[18] I. Dumitriu and S. Pal, Sparse regular random graphs: spectral density and eigenvectors. *Ann. Probab.* 40 (2012), no. 5, 2197–2235.

[19] P. Erdős, On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** (1945), 898–902.

[20] J. Friedman, J. Kahn, and E. Szemerdi, On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, STOC '89, pages 587-598, New York, NY, USA, 1989. ACM.

[21] A. Guionnet, M. Krishnapur and O. Zeitouni, The single ring theorem. *Ann. of Math. (2)* **174** (2011), no. 2, 1189–1217.

[22] J. Kahn, J. Komlós, and E. Szemerédi, On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.* **8** (1995), no. 1, 223–240.

[23] J. Komlós, On the determinant of $(0,1)$ matrices. *Studia Sci. Math. Hungar.* **2** (1967) 7–22.

[24] J. Komlós, On the determinant of random matrices. *Studia Sci. Math. Hungar.* **3** (1968) 387–399.

[25] J. Komlós, Circulated manuscript, 1977. edited version available online at: `http://www.math.rutgers.edu/~komlos/01short.pdf`

[26] M. Krivelevich, B. Sudakov, V. Vu and N. Wormald, Random regular graphs of high degree. *Random Structures Algorithms* **18** (2001), no. 4, 346–363.

[27] M. Ledoux, *The concentration of measure phenomenon. Mathematical surveys and monographs* **89**. Amer. Math. Soc., Providence, RI.

[28] J. E. Littlewood and A. C. Offord, On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.* **12(54)** (1943), p. 277–286.

[29] V. A. Marchenko and L. A. Pastur, The distribution of eigenvalues in certain sets of random matrices, *Mat. Sb.* **72** (1967), 507–536.
[30] B. Maurey, Constructions de suites symétriques. *C.R. Acad. Sci. Paris Sér. A-B* **288** (1979), no. 14, A679–A681.
[31] B. D. McKay, The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.* **40** (1981), p. 203–216.
[32] B.D. McKay, Subgraphs of random graphs with specified degrees. *Congressus Numerantium*, **33** (1981), 213-223.
[33] H. Nguyen, On the singularity of random combinatorial matrices. *SIAM J. Discrete Math.* **27** (2013), no. 1, 447–458.
[34] H. Nguyen, Random doubly stochastic matrices: the circular law. *To appear in Ann. Probab.*
[35] H. Nguyen and V. Vu, Optimal inverse Littlewood-Offord theorems. *Adv. Math.* **226** (2011), no. 6, 5298–5319.
[36] H. Nguyen and V. Vu, Circular law for random discrete matrices of given row sum. *J. Comb.* **4** (2013), no. 1, 1–30.
[37] A. Odlyzko, On subspaces spanned by random selections of ± 1 vectors, *J. Combin. Theory Ser. A.* **47** (1988), no. 1, 124–133.
[38] M. Rudelson, Lower estimates for the singular values of random matrices, *C. R. Math. Acad. Sci. Paris* **342** (2006), 247–252.
[39] M. Rudelson and R. Vershynin, The Littlewood-Offord Problem and invertibility of random matrices. *Advances in Mathematics* **218** (2008), 600–633.
[40] M. Rudelson and R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$. *C. R. Math. Acad. Sci. Paris* **346** (2008), no. 15-16, 893–896.
[41] M. Rudelson and R. Vershynin, Invertibility of random matrices: unitary and orthogonal perturbations. *To appear in J. Amer. Math. Soc.*
[42] G. Schechtman, Lévy type inequality for a class of metric spaces. Martingale Theory and Harmonic Analysis and Banach Spaces, Cleveland 1981. Lecture Notes in Math. 939, 211-215 (1981). Springer.
[43] R. P. Stanley, Enumerative combinatorics. Vol. 2. With a foreword by Gian-Carlo Rota and appendix by Sergey Fomin. Cambridge Studies in Advanced Mathematics, 62. *Cambridge University Press, Cambridge*, 1999.
[44] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 583–602. Univ. California Press, Berkeley, Calif., 1972.*
[45] T. Tao, Topics in random matrix theory. Graduate Studies in Mathematics, 132. *American Mathematical Society, Providence, RI,* 2012.
[46] T. Tao, Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields* **155** (2013), no. 1-2, 231–263.
[47] T. Tao and V. Vu, On random ± 1 matrices: singularity and determinant. *Random Structures Algorithms* **28** (2006), no. 1, 1–23.
[48] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.* **20** (2007), no. 3, 603–628.
[49] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Math.* **169** (2009), 595–632
[50] T. Tao and V. Vu, Random Matrices: The circular Law. *Commun. Contemp. Math.* 10 (2008), no. 2, 261–307.
[51] T. Tao and V. Vu, Random matrices: universality of ESDs and the circular law. *Ann. Probab.* **38** (2010), no. 5, p. 2023–2065, With an appendix by Manjunath Krishnapur.
[52] T. Tao and V. Vu, Additive Combinatorics. Cambridge University Press, 2006.
[53] L. Tran, Local law for eigenvalues of random regular bipartite graphs. *Available at arXiv:math/1310.4606.*
[54] L. Tran, V. Vu, and K. Wang, Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms* **42** (2013), no. 1, 110–134.
[55] V. Vu, Discrete random matrices. *Available at arXiv:math/0611321.*
[56] P.M. Wood, Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.* 22 (2012), no. 3, 1266–1300.

[57] N.C. Wormald, Models of random regular graphs. *Surveys in combinatorics, 1999 (Canterbury),* 239–298, London Math. Soc. Lecture Note Ser., 267, *Cambridge Univ. Press, Cambridge,* 1999.

[58] http://mathoverflow.net/questions/86118/non-enumerative-proof-that-there-are-many-derangements

Department of Mathematics, UCLA, Los Angeles, CA 90095-1555

E-mail address: nickcook@math.ucla.edu